TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

EINAR HILLE C. C. MAC DUFFEE OSCAR ZARISKI

VOLUME 51
JANUARY TO JUNE, 1942

PUBLISHED BY THE SOCIETY
MENASHA, WIS., AND NEW YORK
1942

Chester C. Corbin Library Fund
Bound July, 1942

Composed, Printed and Bound by
The Collegiate Press
George Banta Publishing Company
Menasha, Wisconsin

TABLE OF CONTENTS

VOLUME 51, JANUARY TO JUNE, 1942

SEGMENTS OF ORDERED SETS(1)

BY

W. D. DUTHIE

SECTION I. INTRODUCTION

A set P of elements $a, b, c, \cdots$ in which is defined a binary relation "$\leqq$" having the properties (1) $a \leqq a$ for all $a \in P$, (2) $a \leqq b$ and $b \leqq a$ imply $a = b$, (3) $a \leqq b$, $b \leqq c$ imply $a \leqq c$, will be called an *ordered set*. If for any pair of elements $a, b \in P$ one of the three relations $a \leqq b$, $b \leqq a$, $a = b$ necessarily holds, P is said to be completely ordered(2). In such a set, the subset of all elements x which satisfy the condition $a \leqq x \leqq b$ is the segment joining a and b; however, this definition cannot be used for any pair of elements of an arbitrary ordered set, since it implies that $a \leqq b$. In Section II a definition of a segment is given which is applicable to any pair of elements of an ordered set. Lattices and their fundamental properties can then be defined entirely in terms of the notion of segment. In Section III, the lattice of segments of a lattice is defined, generalizing the lattice of quotient lattices of Ore [3]. A significant difference between our theory and Ore's is that our ordering relation for segments preserves the relation of set-theoretic inclusion between the segments as sets.

Section IV contains a discussion of the convex subsets of a lattice which may be defined in the natural way in terms of segments. An imbedding of any lattice in a complete lattice is obtained by use of the lattice of its convex subsets. In Section V we define in terms of segments two new types of lattice with interesting geometric interpretations.

The following definitions and notation are used: A *lattice* L is an ordered set which contains with any pair of elements a and b their *least upper bound* $a+b$ (an element such that $a+b \geqq a, b$ and $c \geqq a, b$ implies $c \geqq a+b$) and their *greatest lower bound* ab (an element such that $ab \leqq a, b$ and $d \leqq a, b$ implies $d \leqq ab$). An element $0 \in L$ such that $0 \leqq a$ for all $a \in L$ is the *zero* of L; dually an element $1 \in L$ such that $1 \geqq a$ for all $a \in L$ is the *unit* of L. Two elements a and b such that $ab = 0$ are called μ-independent; if $a + b = 1$, they are α-independent(3); and if they are both α-independent, and μ-independent, they are

Presented to the Society, January 1, 1941; received by the editors June 11, 1940.

(1) This paper is the major portion of the author's doctoral thesis, Duthie [9]. (Numbers in brackets refer to the bibliography at the end of the paper.) The omitted portion consists of a fuller treatment of Section II and a discussion of segments and their applications in Boolean algebras.

(2) This terminology differs from the usual designation of *partially ordered* and *linearly ordered*, respectively, for the set P; since this paper is concerned only incidentally with the latter type of order, it is more convenient to make specific mention of it when it occurs.

(3) The α and μ abbreviations for *additive* and *multiplicative* follow a convention of Stone [5], where they are used to denote additive and multiplicative ideals.

complements. A lattice in which every element has a complement is called *complemented*. L is *modular* if $a \leqq c$ implies $(a+b)c = a+bc$ for all $b, c \in L$; *distributive* if $(a+b)c = ac+bc$ for all $a, b, c \in L$; *complete* if any subset of elements of L has a least upper and greatest lower bound in L. Two lattices L and L^* are *isomorphic* if there is a one-one mapping of their elements such that $a \leftrightarrow a^*$, $b \leftrightarrow b^*$ implies $a+b \leftrightarrow a^*+b^*$ and $ab \leftrightarrow a^*b^*$; they are *dually isomorphic* if $a+b \leftrightarrow a^*b^*$ and $ab \leftrightarrow a^*+b^*$. The set-theoretical intersection of two subsets A and B of elements of a lattice will be denoted by $A \cap B$. $a \prec b$ means $a<b$ and $a<x<b$ has no solution.

SECTION II. LATTICE PROPERTIES IN TERMS OF SEGMENTS

1. **Definition of a segment and a lattice.** Let a, b be any two elements of an ordered set P. If $a \leqq b$, "the segment $[a, b]$ joining a and b" is defined to be the set of all elements x of P such that $a \leqq x \leqq b$. We order such segments so as to preserve set-theoretic inclusion; that is, $[a, b] \leqq [c, d]$ if and only if $a \geqq c$ and $b \leqq d$ and hence if and only if the set $[c, d]$ contains the set $[a, b]$.

If neither $a \leqq b$ nor $b \leqq a$, we define the symbol $[a, b]$ as the set-theoretic intersection of all segments $[u, v]$ containing both a and b. It is easily seen that $[a, b]$ will be a segment if and only if a and b have a least upper bound $a+b$ and a greatest lower bound ab in P, and in this case, it equals $[ab, a+b]$. We thus arrive at the following definition:

DEFINITION 2.1. *The segment joining any pair of elements x and y of an ordered set P is the set of all elements $z \in P$ which satisfy the condition*

$$xy \leqq z \leqq x + y.$$

This set will be denoted by the symbol $[x, y]$. The elements xy and $x+y$ will be called the lower and upper extremities, respectively, of the segment $[x, y]$.

COROLLARY 2.2. *An ordered set P is a lattice if and only if there is a segment $[x, y]$ joining every pair of elements $x, y \in P$.*

Because of the uniqueness of least upper and greatest lower bounds in a lattice L, the following corollaries are also immediate consequences of Definition 2.1.

COROLLARY 2.3. *There is one and only one segment joining each pair of elements of a lattice L.*

COROLLARY 2.4. *Two segments coincide if and only if their extremities coincide.*

Since the condition $xy \leqq z \leqq x+y$ is transformed into itself by dualization, being a segment is a self-dual property of subsets of an ordered set.

2. **Principal ideals as segments.** As an example of a segment in a lattice L, consider the principal μ-ideal or α-ideal generated by an element $a \in L$, which

is denoted by $(a)_\mu$ or by $(a)_\alpha$ respectively. Then $(a)_\mu$ ($(a)_\alpha$) is the set of all elements of the form xa ($x+a$), $x\in L$; but that is the same as the set of all $y\in L$ such that $y\leqq a$ ($y\geqq a$). Hence if L has a zero element (unit element), then $(a)_\mu=[0, a]$ ($(a)_\alpha=[0, 1]$) proving

THEOREM 2.5. *If L is a lattice with a zero (unit), then a segment of L is a principal μ-ideal (α-ideal) if and only if it contains the zero (unit) element.*

This theorem is used in what follows as the definition of principal ideals, and other concepts of lattice theory will also be characterized by means of segments; hence it is desirable to show that all the fundamental definitions of the theory can be stated in terms of segments, and the remainder of this section is devoted to that purpose.

3. **Modularity.** Given a segment and a point in a completely ordered set such as the real line, it is trivial that if the segments joining the point to the extremities of the given segment coincide, then the segment is a point. On the other hand, in the case of segments and elements of a lattice, this fact is not only not trivial but even untrue, as is shown by

THEOREM 2.6. *A lattice L is modular if and only if the identity of the segments joining an element to the extremities of the segment joining two other elements implies the identity of these two elements.*

Proof. It is known (Ore [3, p. 413]) that L is modular if and only if for any a, b, c, $a\geqq b$ and $a+c=b+c$, $ac=bc$ imply $a=b$. Thus by Corollary 2.4, L is modular if and only if $a\geqq b$ and $[a, c]=[b, c]$ imply $a=b$. Since $a+b=ab$ if and only if $a=b$, and since $a+b\geqq ab$ always, L is modular if and only if for any a, b, c $[a+b, c]=[ab, c]$ implies $a=b$. This last statement is equivalent to the theorem.

4. **Distributivity.** Since the definition of distributivity to be given below involves set-theoretic intersection of segments, the following simple lemma will be needed.

LEMMA 2.7. $[a, b]\cap[c, d]=[ab+cd, (a+b)(c+d)]$ *if the intersection is non-empty.*

Proof. Any element $x\in L$ which belongs to both $[a, b]$ and $[c, d]$ is subjected to the simultaneous conditions $ab\leqq x\leqq a+b$ and $cd\leqq x\leqq c+d$. Hence the smallest x which satisfies these conditions is $x=ab+cd$ and the largest is $x=(a+b)(c+d)$. The intersection will be non-empty if and only if

$$ab + cd \leqq (a + b)(c + d).$$

In a completely ordered set such as the real line, a point lying on the segment joining two points is uniquely represented as the intersection of the segments joining it to each of these points. That this is likewise a non-trivial property of elements and segments of a lattice is shown by

THEOREM 2.8. *A lattice is distributive if and only if any element belonging to the segment joining two elements a, b is the intersection of the segments joining it to each element a, b.*

Proof(4). If $c\in[a, b]$, then by Lemma 2.7, $[a, c]\cap[c, b]=[ac+bc, (a+c)(b+c)]$. Hence it suffices to show that L is distributive if and only if

(i) $a+b\geqq c\geqq ab$ implies $ac+bc=c$ and $(a+c)(b+c)=c$.

Now (i) is easily seen to be equivalent to

(ii) L contains no non-distributive sublattice of order five.

But (ii) is a necessary and sufficient condition for distributivity, (G. Birkhoff [2, Theorem 3.7, Theorem 5.2 with second corollary]).

5. **Complementation.** The fact that an element a of a lattice has a complement is easily translated into segment terms by the following definition.

DEFINITION 2.9. *An element a of a lattice L with zero and unit elements is a complement of another element b if and only if the segment joining a and b is the whole lattice L.*

The following proof of the well known fact that complements in a distributive lattice are unique illustrates the use of segments in deriving other lattice properties: Assume that an element a has two complements, a' and a''. Then $[a'', a']=[a'', a']\cap[0, 1]=[a'', a']\cap[a', a]=[a', a']$, where the last equality follows from Theorem 2.8. Hence by Corollary 2.4, $a''+a'=a'=a''a'$, and so $a''=a'$.

SECTION III. FORMAL PROPERTIES OF SEGMENTS

1. **Segments as lattices.** As an immediate corollary of Definition 2.1, we have the following properties of the symbol $[a, b]$:

$$[a, b] = [b, a] = [ab, a + b] = [a + b, ab].$$

Thus there is no loss in generality if the symbol $[a, b]$ is considered to be an *ordered pair* with the additional condition $a\leqq b$; then $[a, b]$ is a sublattice of L whose zero element is a and whose unit element is b. Such a sublattice has been called by Ore [3] a "quotient structure."

Properties of L such as modularity, distributivity, completeness, and complementation carry over from L to $[a, b]$(5).

THEOREM 3.1. *A sublattice of the direct sum of two lattices is a segment if and only if it is the direct sum of a segment in one by a segment in the other.*

Proof. Let (a, b) and (c, d) be the lower and upper extremities of a segment in the direct sum. Then an arbitrary element of the segment satisfies the con-

(4) A longer, but purely combinatorial, proof of this theorem due in part to J. von Neumann is given in Duthie [9]. This proof was supplied by the referee, to whom the author is indebted for considerable revision of this section.

(5) The proofs are given in von Neumann [6, vol. 1].

dition $(a, b) \leqq (x, y) \leqq (c, d)$. But from the definition of direct sum, this is equivalent to the two conditions $a \leqq x \leqq c$ and $b \leqq y \leqq d$ together.

This theorem is used in the following example.

EXAMPLE 3.2. Let E_1 be the completely ordered real line, and consider the real plane as the (lattice) direct sum $E_1 \oplus E_1$. With respect to this ordering, segments in the plane are of three types: (1) single points, (2) closed segments parallel to either axis, (3) closed rectangles with sides parallel to the axes.

The usual theorems on the modularity, distributivity, and complementation of the direct sum of two lattices are corollaries of Theorem 3.1, in view of the formulation of these properties given in Section II.

2. **Ordered sets of segments.** The set of segments of a lattice, which will be denoted by L_s, may be ordered as follows:

DEFINITION 3.3. *There is an element* $\phi \in L_s$ *such that for all* $[a, b] \in L_s$, $\phi < [a, b]$.

The element ϕ should not be confused with the element $[0, 0]$, in case L has a zero element.

DEFINITION 3.4. $[a, b] \leqq [c, d]$ *if and only if* $a \geqq c$ *and* $b \leqq d$. *The segment* $[c, d]$ *is called an extension of* $[a, b]$.

It is immediately apparent from Definition 3.4 that the set $L_s - \phi$ is closely related to the direct sum of L_d (the dual of L obtained by interchanging $\leqq$ and $\geqq$ in L) and L. In fact, the only thing which prevents $L_s - \phi$ and $L_d \oplus L$ from being isomorphic is the restriction $a \leqq b$ on the elements $[a, b]$ of L_s. We are thus led to the consideration of a subset S of elements (a, b) of $L_d \oplus L$ subjected to a similar restriction. Since $(a, b) + (c, d) = (ac, b+d)$ and $ac \leqq b+d$ if $a \leqq b$ and $c \leqq d$, the set S is closed under addition. However, it is not closed under multiplication, for $(a, b) \cdot (c, d) = (a+c, bd)$ and the condition $a+c \leqq bd$ will not in general hold. Hence S is not a *sublattice* of $L_d \oplus L$, but it is possible to make it into a *lattice* by adjoining an element θ with the properties

$$\left.\begin{aligned} \theta \cdot (a, b) &= (a, b) \cdot \theta = \theta, \\ \theta + (a, b) &= (a, b) + \theta = (a, b), \end{aligned}\right\} \qquad \text{for all } (a, b) \in S,$$

$$(a, b) \cdot (c, d) = \theta \qquad \text{if } a + c \nless bd, \text{ and } a + c \neq bd.$$

This discussion may then be summarized as follows:

THEOREM 3.5. *The set of segments of a lattice L, ordered as in Definitions* 3.3 *and* 3.4, *is a lattice isomorphic to the subset S of* $L_d \oplus L$ *with the element* θ *adjoined, and the elements* ϕ *and* θ *correspond under the isomorphism.*

Furthermore, the operation of multiplication in L_s is equivalent to set-theoretical intersection of segments, by virtue of Lemma 2.7.

If the sum of two segments is defined as the intersection of all segments containing them, then the same L_s is obtained as above. However, the derivation by use of the direct sum of L and L_d makes some later results more easily proved.

3. **Relations between** L **and** L_s. In looking for special properties of L_s, it is natural to seek first those properties of L which carry over to L_s. By definition, L_s always has a zero element, even if L does not, and it is easily seen that L_s will have a unit if and only if L has both a zero and a unit, in which case the unit is $[0, 1]$. It is also obvious from the definition of the lattice operations in L_s that if L is complete, then L_s is also complete. Another property of L_s which carries over from L is given in

THEOREM 3.6. *If L is complemented, then L_s is complemented.*

Proof. Two elements $[a, b]$ and $[c, d]$ of L_s are complements if and only if

$$[a, b]\cdot[c, d] = \phi = [a + c, bd],$$
$$[a, b] + [c, d] = [0, 1] = [ac, b + d].$$

Hence $[a, b]$ and $[c, d]$ will be complements if their lower extremities are μ-independent in L and their upper extremities are α-independent in L, and $a+c \nless bd$, and $a+c \neq bd$, conditions which can obviously be satisfied if L is complemented.

This proof also shows that an element of L_s may have more than one complement, even though complementation in L is unique; hence there is no hope that distributivity will carry over from L to L_s in general. That it does not even carry over into modularity is shown by

EXAMPLE 3.7. Let L be the completely ordered set $0 < a < 1$, and consider the elements $[0, 0]$, $[0, a]$, and $[1, 1]$ of L_s. Since $[0, 0] < [0, a]$, the modular law may be applied to these three elements, giving

$$\{[0, 0] + [1, 1]\}[0, a] = [0, 1]\cdot[0, a] = [0, a];$$

but

$$[0, 0] + [1, 1]\cdot[0, a] = [0, 0] + \phi = [0, 0].$$

Hence L_s is not modular. Incidentally, L is not complemented, and neither is L_s.

Two more general types of lattices have recently been defined and discussed by Wilcox [8, pp. 495–496]. Making use of a binary relation $(b, c)M$, which means $(a+b)c = a+bc$ for all $a \leq c$, a lattice is said to be *symmetric* if it satisfies

AXIOM A. *If $bc = 0$ and $(b, c)M$, then $(c, b)M$.*

L is called *semi-modular* if, in addition, it satisfies

AXIOM B. *If $bc \neq 0$, then $(b, c)M$.*

THEOREM 3.8. *If L is modular and contains more than two elements, then L_s satisfies Axiom* B *but not Axiom* A.

Proof. In Example 3.7, it was shown that $([1, 1], [0, a])M$ does not hold; but $[0, a]\cdot[1, 1]=\phi$ and $([0, a], [1, 1])M$ since $[1, 1]$ is atomic(6). If L has only two elements, then L_s has the four elements $[0, 1]$, $[1, 1]$, $[0, 0]$, ϕ and is a Boolean algebra—hence *a fortiori* satisfies Axiom A.

Now consider the so-called one-sided modular law, valid in any lattice whatsoever, written in terms of elements of L_s: $[a_1, a_2] \leqq [c_1, c_2]$ implies

$$\{[a_1, a_2] + [b_1, b_2]\}[c_1, c_2] \geqq [a_1, a_2] + [b_1, b_2][c_1, c_2].$$

Assuming that $[b_1, b_2]\cdot[c_1, c_2] \neq \phi$ and $[a_1, a_2] \neq \phi$ (the equality of the two sides of the above expression is trivial if $[a_1, a_2]=\phi$) all elements and combinations of elements in the inequality are different from ϕ, and hence the rules for combining elements in L_s will be valid. Therefore

$$\{[a_1, a_2] + [b_1, b_2]\}[c_1, c_2] = [a_1b_1, a_2 + b_2][c_1, c_2] = [a_1b_1 + c_1, (a_2 + b_2)c_2].$$

Since L is modular, and $c_1 \leqq a_1 \leqq a_2 \leqq c_2$,

$$\begin{aligned}[a_1b_1 + c_1, (a_2 + b_2)c_2] &= [a_1(b_1 + c_1), a_2 + b_2c_2] = [a_1, a_2] + [(b_1 + c_1), b_2c_2] \\ &= [a_1, a_2] + [b_1, b_2][c_1, c_2];\end{aligned}$$

proving that Axiom B holds in L_s.

L_s thus belongs to a class of lattices satisfying Axiom B but not Axiom A. Since the term *semi-modular* has already been applied to those lattices satisfying both axioms, the larger class satisfying Axiom B will be called *pseudo-modular*. This property and the analogous one of *pseudo-distributivity* will be discussed in Section V.

4. **Special properties of L_s.** It has already been noted that L_s has a zero even if L does not. L_s has other properties not necessarily shared by L, some of which will be exhibited in the theorems of this paragraph.

THEOREM 3.9. *There is a one-to-one correspondence between elements of L and atomic elements of L_s, and every non-atomic element of L_s can be expressed uniquely as the sum of exactly two atomic elements.*

Proof. The correspondence is $a \leftrightarrow [a, a]$; furthermore, $[a, a]+[b, b] = [ab, a+b]=[a, b]$ if $a \leqq b$, and $[a, a]\cdot[b, b]=\phi$ if $a \neq b$.

THEOREM 3.10. *The principal α-ideal generated by an element $[a, b]$ of L_s is isomorphic to the direct sum of the principal μ-ideal generated by its lower extremity and the principal α-ideal generated by its upper extremity.*

(6) An element p of a lattice L with a zero is atomic if $p \succ 0$.

Proof. $([a, b])_\alpha$ is composed of all $[x, y]$ such that $x \leqq a$ and $y \geqq b$; hence from the definition of direct sum, it is isomorphic to $(a)_\mu \oplus (b)_\alpha$.

COROLLARY 3.11. *The set of all principal μ-ideals (α-ideals) of a lattice L with a zero (unit) element is a divisorless*(7) *principal α-ideal of L_s, which is isomorphic (dually isomorphic) to L.*

Proof. Theorems 3.10 and 2.5. The ideal is divisorless because $[0, 0]$ (or $[1, 1]$) is an atomic element.

COROLLARY 3.12. *Every principal α-ideal of L_s can be represented uniquely as the intersection of two divisorless principal α-ideals.*

Proof. Theorem 3.9.

5. **Representation of segments by principal ideals.** Since $(a)_\alpha$ is the set of all $x \in L$ such that $x \geqq a$ and $(b)_\mu$ is the set of all $x \in L$ such that $x \leqq b$, it is possible to reverse the procedure of §2 of Section II and express segments in terms of principal ideals by

THEOREM 3.13. *Every segment $[a, b]$ is uniquely represented by the set-theoretical intersection of the principal α-ideal generated by its lower extremity and the principal μ-ideal generated by its upper extremity.*

Lemma 2.7 then becomes a corollary of this theorem, since

$$\begin{aligned}[a, b] \cap [c, d] &= \{(a)_\alpha \cap (b)_\mu\} \cap \{(c)_\alpha \cap (d)_\mu\} \\ &= \{(a)_\alpha \cap (c)_\alpha\} \cap \{(b)_\mu \cap (d)_\mu\},\end{aligned}$$

which by Corollary 3.11

$$\begin{aligned} &= (a + c)_\alpha \cap (bd)_\mu \\ &= [a + c, bd].\end{aligned}$$

SECTION IV. CONVEX SUBSETS OF A LATTICE

1. **Convexity.** Segments are used to define the convex subsets of a lattice in the same manner as in metric or linear spaces.

DEFINITION 4.1. *A subset S of elements of a lattice L is called convex if it contains the segment joining any pair of its elements.*

Since $a, b \in S$ implies $[a, b] = [ab, a+b] \subset S$, we have the obvious

COROLLARY 4.2. *A convex subset of L is a sublattice of L, and if it is not a segment it will lack either a zero or unit or both*(8).

(7) A μ-ideal (α-ideal) is divisorless if it is contained in no other μ-ideal (α-ideal) except L itself.

(8) This corollary shows that "convex subset" as here defined is equivalent to what Ore [3] calls "dense substructure." It is not however equivalent to the definition of convexity given by Birkhoff [2], since there the convex subsets need not be lattices.

COROLLARY 4.3. *If a lattice L has finite dimension, then all convex sets are segments.*

2. **Ideals and convex sets.** In the two preceding sections, the relationship between segments and principal ideals was discussed. Analogous relationships will now be derived for convex sets and ideals.

THEOREM 4.4. *The intersection of an α-ideal and a μ-ideal is a convex set and every convex set is the intersection of the least α- and μ-ideals containing it,*

Proof. Let C be the intersection of an α-ideal A and a μ-ideal M. Since A and M are sublattices of L, C is also a sublattice, and if $a, b \in C$, then $ab, a+b \in c$. Now M contains $a+b$ and therefore all $x \in L$ such that $x \leqq a+b$; dually, A contains ab and with it all $y \in L$ such that $ab \leqq y$. Hence $A \cap M$ contains all $x \in L$ such that $ab \leqq x \leqq a+b$.

Now assume that A and M are the least α- and μ-ideals containing C. Then M (A) is composed of elements $x \in C$ and all elements $v \in L$ such that $v \leqq x$ $(v \geqq x)$ for any $x \in C$. Hence $C = A \cap M$.

COROLLARY 4.5. *All ideals are convex sets.*

COROLLARY 4.6. *Every convex set is the intersection of all the ideals containing it.*

3. **The lattice of convex subsets of** L. The set-theoretical intersection of a set of convex sets is again convex and so the set of convex subsets of a lattice L can be made into a complete lattice, which will be denoted by L_{cs}, by the usual process of defining the lattice sum of any number of convex sets as the intersection of all convex sets containing each of the sets. The symbols $\wedge$ and $\vee$ will be used to denote the lattice operations in L_{cs}. The zero and unit elements of L_{cs} are the empty set and L itself.

Most of the significance of this lattice lies in

THEOREM 4.7. *L_s is a sublattice, but not in general a complete sublattice, of L_{cs}.*

Proof. It has already been pointed out that L_s is not necessarily complete. If $A = [a, b]$ and $B = [c, d]$, then $A \wedge B = [a, b] \cdot [c, d]$ by Lemma 2.7. Likewise, $A \vee B = [a, b] + [c, d]$, since any convex set containing both A and B must contain the segments joining the elements ab, cd and $a+b$, $c+d$, hence $abcd$ and $a+b+c+d$.

In particular, L_{cs} has as atomic elements the atomic elements of L_s (although they no longer form a basis for L_{cs}) and hence a class of divisorless principal α-ideals each of which contains as a sublattice the corresponding divisorless principal α-ideal of L_s. Hence if L has a zero or unit element, then by Corollary 3.11 we have an imbedding of a lattice isomorphic or dually iso-

morphic to L in a complete lattice. Moreover, this lattice is isomorphic to the lattice of μ-ideals or α-ideals of L, by virtue of

THEOREM 4.8. *If L is a lattice with a zero (unit) element, then a convex subset A of L is a μ-ideal (α-ideal) of L if and only if it contains the zero (unit) element of L.*

Proof. A being convex, it must contain with each of its elements a the segment $[0, a]$ ($[a, 1]$). Hence with each a it contains all elements of the form xa ($x+a$), $x \in L$, and with each $a, b \in A$, $a+b$ (ab). The converse is obvious.

In the case of distributive lattices, the lattice of μ- or α-ideals is known to be distributive; hence $([0, 0])_\alpha$ and $([1, 1])_\alpha$ (in $L_{\alpha s}$) are also distributive, and the above imbedding process preserves this lattice property.

If L is a Boolean algebra, then by introducing an operation called orthocomplementation, it is possible to identify ideals in such a way as to make this operation actually a complementation (at the same time preserving the completeness and distributivity of the lattice of ideals). Hence one obtains an imbedding of a Boolean algebra in a complete Boolean algebra(9).

SECTION V. PSEUDO-MODULAR AND PSEUDO-DISTRIBUTIVE LATTICES

1. **Definitions and examples.** In Section III, it was shown that the L_s of a modular lattice satisfied the condition of Axiom B, and such a lattice was called *pseudo-modular*. The following corollary is an immediate consequence of Axiom B and gives an equivalent formulation of the property of pseudo-modularity which is more conveniently specialized to the corresponding notion of pseudo-distributivity.

COROLLARY 5.1. *If every sublattice of a lattice L with a zero element which does not contain the zero element is modular, then L is pseudo-modular, and conversely.*

The method of defining the analogous property of pseudo-distributivity is then obvious:

DEFINITION 5.2. *A lattice with zero element is said to be pseudo-distributive if and only if every sublattice not containing the zero element is distributive.*

It is now possible to make a formal distinction between the lattices of segments of modular and distributive lattices by

THEOREM 5.3. *The lattice of segments of a modular (distributive) lattice is pseudo-modular (pseudo-distributive).*

(The theorem has already been proved for modularity, but the following proof is applicable to both cases.)

(9) A detailed discussion of the lattices of ideals of Boolean algebras and distributive lattices will be found in Stone [4] and [5].

Proof. Every sublattice of an L_s which does not contain the zero of L_s is also a sublattice of one of the divisorless principal α-ideals generated by an atomic element of L_s. Hence by Theorem 3.10 it is isomorphic to a sublattice of the direct sum $(a)_\mu \oplus (a)_\alpha$, which is modular (distributive) if L is modular (distributive).

This theorem completes the discussion of lattices of segments, and the remainder of this section will be devoted to the development of properties of pseudo-modularity and pseudo-distributivity, which seem to have some interest in themselves, especially the latter, since it is applicable to modular lattices.

Examples of pseudo-modular lattices which are not lattices of segments of some modular lattices are the semi-modular lattices, one type of which can be constructed from modular lattices by a process due to Wilcox [8, p. 497 ff.]. Pseudo-distributive lattices may be constructed from distributive lattices by the same process, the restrictions on the distributive lattices being less stringent since there is no need to satisfy Axiom A of Section III. The method is as follows:

Let D be an arbitrary distributive lattice with zero element. If S is a μ-ideal of D with 0 deleted, then the set $L = D - S$ is a lattice with its operations $\cup, \cap$ defined by

$$a \cup b = a + b,$$

$$a \cap b = \begin{cases} ab & \text{if} \quad ab \in L, \\ 0 & \text{if} \quad ab \in S. \end{cases}$$

The proof that $a \cup b$ and $a \cap b$ are actually effective as least upper and greatest lower bounds in L depends only on the fact that S is a μ-ideal with 0 deleted, and not on the distributivity of D, so the argument given by Wilcox applies directly. The distributivity of any sublattice of L not containing 0 is immediate from the definitions of $a \cup b$ and $a \cap b$ and the distributivity of sublattices of D.

2. **Duality considerations.** The above examples show immediately that pseudo-modularity and pseudo-distributivity, unlike modularity and distributivity, are neither self-dual properties of a lattice nor do they imply their duals. A repetition of the above construction using an α-ideal in place of the μ-ideal will yield examples of *dual pseudo-modular* or *dual pseudo-distributive* lattices from modular or distributive lattices with unit elements. However, under certain conditions pseudo-distributivity is equivalent to dual pseudo-distributivity. We first give examples to show that modularity and pseudo-distributivity are independent properties of a lattice.

EXAMPLE 5.4. L_5 is a 5-element lattice with order relations as follows: $0 < a < 1$, $0 < b < c < 1$, $a \not< c$.

EXAMPLE 5.5. Define order in a 6-element lattice L_6 by $0<a_i<b<1$, $i=1, 2, 3$.

L_5 is a non-modular pseudo-distributive lattice, while L_6 is modular and pseudo-distributive.

THEOREM 5.6. *In a complete complemented modular lattice, pseudo-distributivity is equivalent to dual pseudo-distributivity.*

To prove this theorem, the following lemma is needed:

LEMMA 5.7. *In a modular lattice L, the segment $[ab, a]$ is isomorphic to the segment $[b, a+b]$ for all $a, b \in L$.*

The proof of this lemma is given in Birkhoff [1] and in Ore [3, p. 418, Theorem 2].

Proof of the theorem. By duality, we need only prove that pseudo-distributivity implies dual pseudo-distributivity. Assume that L is not dual pseudo-distributive; then there is some sublattice S of L which does not contain the unit of L and which is not distributive. Since L is complete, S has a unit element $x\neq 1$, and S is a sublattice of $[0, x]$; also L is complemented, so there exists an element $y\in L$ $(\neq 0)$ such that $x+y=1$, $xy=0$. Then by Lemma 5.7, $[0, x]$ is isomorphic to $[y, 1]$, which is distributive if L is pseudo-distributive, contradicting the assumption that S is not distributive.

Remark. The hypotheses of this theorem can be weakened somewhat; for instance, modularity may be replaced by the condition $[ab, a]$ isomorphic to $[b, a+b]$ for all $a, b\in L$. Ward [7, p. 448] has shown that such a lattice need not be modular, but is modular if one of the chain conditions holds. Example 5.5 shows that it is not possible to omit the requirement of complementation, since L_6 is pseudo-distributive but not dual pseudo-distributive.

3. ***m*-distributivity in modular lattices.** In lattices on which there is defined a positive monotone dimension or measure function(10) $d(a)$ it is possible to make a slight generalization of the notions of pseudo-distributivity and its dual as follows:

DEFINITION 5.8. *A lattice L with a dimension function $d(x)$ is said to be m-distributive (dual m-distributive) if all sublattices whose zero elements (unit elements) have dimension $d(a)\geqq m$ $(d(a)\geqq (n-m)$ in the finite case and $d(a)\geqq (1-m)$ in the continuous case) are distributive.*

COROLLARY 5.9. *In a lattice in which $d(x)$ has the range $0, 1, 2, \cdots, n$, 1-distributivity (dual 1-distributivity) is equivalent to pseudo-distributivity (dual pseudo-distributivity).*

(10) The function $d(a)$ is a numerical function whose range is either discrete $(0, 1, 2, \cdots, n)$ or continuous $(0\leqq d(a)\leqq 1)$, and has the properties (1) $a\leqq b$ implies $d(a)\leqq d(b)$, (2) $d(a)+d(b)=d(a+b)+d(ab)$ for all $a, b\in L$. For more details, see von Neumann [6, vol. 1], or Birkhoff [2, chap. 4].

When $d(x)$ is continuous, it never has a value m such that m-distributivity or its dual is equivalent to pseudo-distributivity or its dual.

COROLLARY 5.10. *Every lattice of dimension n is $(n-1)$-distributive and dual $(n-1)$-distributive. In particular, all lattices of dimension 2 are both pseudo-distributive and dual pseudo-distributive.*

It is obvious that in either the finite or continuous case, 0-distributivity is equivalent to distributivity as is also dual 0-distributivity.

EXAMPLE 5.11. An example of an m-distributive lattice which is not pseudo-distributive is the free modular lattice generated by three elements. It is of finite dimension ($n=8$) and is both 4-distributive and dual 4-distributive.

Since lattices of finite or continuous dimension are complete, a slight modification of the proof of Theorem 5.6 yields

THEOREM 5.12. *In a complemented modular lattice of finite or continuous dimension, m-distributivity is equivalent to dual m-distributivity.*

The 4-distributive lattice of Example 5.11, while complete and modular, is not complemented. The class of m-distributive lattices which are complete, modular, and complemented is greatly restricted by the following,

LEMMA 5.13. *If a complete complemented modular lattice is irreducible, then for any $a, b \in L$, $[a, b]$ is irreducible*(11).

An immediate consequence of this lemma and Corollary 5.10 is

THEOREM 5.14. *An irreducible complemented modular lattice of finite dimension n is m-distributive only if $m=n-1$, and there are no irreducible m-distributive or pseudo-distributive complemented modular lattices of continuous dimension.*

In particular, the only pseudo-distributive projective geometries are one-dimensional, and their pseudo-distributivity is trivial.

Now consider the reducible(12) complemented modular lattices of finite or continuous dimension. The *index* of such a lattice is defined as follows:

DEFINITION 5.15. *If Z is the center of a reducible complemented modular lattice L of finite or continuous dimension, then $m=\max_{z \in Z} \{\min [d(x), d(x')]\}$ (where x' denotes the (unique) complement of x) is the index of L.*

THEOREM 5.16. *A complemented modular lattice of index m which is m-distributive is a Boolean algebra.*

(11) For proof of this lemma, see von Neumann [6, vol. 1].

(12) These lattices are discussed in von Neumann [9, vol. 2, Part III], and Birkhoff [2 chap. 4].

Proof. Let a be an element of Z such that $d(a)=m$. Since $a\in Z$, $L=[0, a]+[0, a']$; also $d(a)=m$ and $d(a')\geqq m$, making both $[0, a]$ and $[0, a']$ distributive by Theorem 5.12. Hence L is distributive and is a Boolean algebra.

COROLLARY 5.17. *The only pseudo-distributive reducible complemented modular lattices of finite or continuous dimension are Boolean algebras.*

The preceding sequence of theorems and corollaries indicates that if significant examples of pseudo- and m-distributive modular lattices and their duals are to be found, the requirement of complementation must be dropped. The lattices of invariant subgroups of a group and ideals of a ring are lattices of this type, and the determination of what, if any, algebraic properties of groups and rings are implied by the various degrees of distributivity of their lattices of invariant subgroups and ideals is a problem which might be of some significance. For instance, consider the following example of the lattice of (invariant) subgroups of the quaternion group ± 1, $\pm i$, $\pm j$, $\pm k$. Its proper invariant subgroups are $A_1=(+1, +i, -1, -i)$, $A_2=(+1, +j, -1, -j)$, $A_3=(+1, +k, -1, -k)$, and $B=(+1, -1)$, and these together with the whole group and the unit element constitute a lattice dually isomorphic to the lattice L_6 of Example 5.5. Hence it is dual pseudo-distributive.

BIBLIOGRAPHY

1. G. Birkhoff, *On combinations of subalgebras*, Proceedings of the Cambridge Philosophical Society, vol. 29 (1933), pp. 441–464.

2. ———, *Lattice Theory*, American Mathematical Society Colloquium Publications, vol. 25, 1940.

3. O. Ore, *On the foundations of abstract algebra* I, Annals of Mathematics, (2) vol. 36 (1935), pp. 406–437.

4. M. H. Stone, *The theory of representations for Boolean algebras*, these Transactions, vol. 40 (1936), pp. 37–111.

5. ———, *Topological representations of distributive lattices and Brouwerian logics*, Časopis pro Pěstování Matematiky a Fysiky, vol. 67 (1937), pp. 1–25.

6. J. von Neumann, *Continuous Geometry*, vols. 1, 2, Planographed Princeton Mathematical Notes.

7. M. Ward, *A characterization of Dedekind structures*, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 448–451.

8. L. R. Wilcox, *Modularity in the theory of lattices*, Annals of Mathematics, (2), vol. 40 (1939), pp. 490–505.

9. W. D. Duthie, *Segments in ordered sets*, doctoral thesis, Princeton University, 1940.

PRINCETON UNIVERSITY,
PRINCETON, N.J.

THEORY OF MONOMIAL GROUPS

BY

OYSTEIN ORE

The process of imbedding a group in a larger group of some prescribed type has been one of the most useful tools in the investigation of properties of groups. The three principal types of representation of groups, each with its particular field of usefulness, are the following:

1. Permutation groups.
2. Monomial groups.
3. Linear or matrix representations of groups.

These three types of representation correspond to an imbedding of the group in the following groups:

1. The symmetric group.
2. The complete monomial group.
3. The full linear group.

The symmetric group and the full linear group have both been exhaustively investigated and many of their principal properties are known. A similar study does not seem to exist for the complete monomial group. Such a general theory seems particularly desirable in view of the numerous recent investigations on finite groups in which the monomial representations are used in one form or another to obtain deep-lying theorems on the properties of such groups. The present paper is an attempt to fill this lacuna.

In this paper the monomial group or symmetry is taken in the most general sense(1) where one considers all permutations of a certain finite number of variables, each variable being multiplied also by some element of a fixed arbitrary group H. In the first chapter the simplest properties such as transformation, normal form, centralizer, etc., are discussed. Some of the auxiliary theorems appear to have independent interest. One finds that the symmetry contains a normal subgroup, the basis group, consisting of all those elements which do not permute the variables. The symmetry splits over the basis group with a group isomorphic to the symmetric group as one representative group. A complete solution of the problem of finding all representative groups in this splitting of the symmetry is given. This result is of interest since it gives a general idea of the solution of the splitting problem in a fairly complicated case.

In the second chapter all normal subgroups of the symmetry are deter-

Presented to the Society, February 22, 1941; received by the editors November 28, 1940.

(1) See W. Specht, *Eine Verallgemeinerung der symmetrischen Gruppe*, Schriften des mathematischen Seminars, Berlin, vol. 1 (1932), also W. Specht, *Eine Verallgemeinerung der Permutationsgruppen*, Mathematische Zeitschrift, vol. 37 (1933), pp. 321–341.

mined. This is, as one knows, a simple problem in the case of a symmetric group. For the monomial group the problem becomes very complicated and it is found that usually a large number of normal subgroups of various types must exist. The solution depends on the determination of certain types of subgroups of direct products. The successive commutator groups of the symmetry are determined as an application[2].

The most difficult problem in this theory is the determination of all automorphisms of the symmetry. The solution given in the third chapter depends on the previous results on the splitting of the symmetry and the form of its normal subgroups. This case is different from the case of the symmetric group, for outer automorphisms will usually occur. The necessary and sufficient condition for all automorphisms to be inner automorphisms follows from the general result. These investigations are considerably complicated by several exceptions which occur in groups of the lowest orders. These cases have also been completely investigated except in the case $m=6$. Here all the necessary preparations for the calculations have been made, but since the actual solution depends on a somewhat laborious method, the work has not been carried through. It would be desirable if someone would complete the theory at this point.

The final chapter is concerned with the imbedding of an arbitrary group G in a monomial group $\Sigma(H)$. Relations between the various representations are discussed and the position of the imbedded group G in Σ_m is investigated. Its centralizer and normalizer are determined, and the automorphisms of G induced by the elements in the normalizer are obtained.

Chapter I. The symmetries

1. **Definitions.** Let H be some group, finite or infinite. A ***monomial substitution*** over H is a linear transformation

$$\rho = \begin{pmatrix} x_1, & x_2, & \cdots, & x_m \\ r_1x_{i_1}, & r_2x_{i_2}, & \cdots, & r_mx_{i_m} \end{pmatrix} \tag{1}$$

where each variable is changed into some other variable multiplied by an element of H. We shall call the elements r_i in (1) the ***factors*** or ***multipliers*** in ρ. The multiplication r_ix_i is a formal one to be taken only as a pair (r_i, x_i) with the associative property $r(sx)=(rs)x$.

If another such monomial substitution

$$\kappa = \begin{pmatrix} x_1, & x_2, & \cdots, & x_m \\ k_1x_{j_1}, & k_2x_{j_2}, & \cdots, & k_mx_{j_m} \end{pmatrix} \tag{2}$$

is given, then the product of ρ and κ is defined by

(2) W. K. Turkin, *Ueber Herstellung und Anwendungen der monomialen Darstellungen endlicher Gruppen*, Mathematische Annalen, vol. 111 (1935), pp. 743–747.

$$\rho\kappa = \begin{pmatrix} x_1, & \cdots, & x_m \\ k_1 r_{j_1} x_{i_{j_1}}, & \cdots, & k_m r_{j_m} x_{i_{j_m}} \end{pmatrix}.$$

The inverse of κ is

$$\kappa^{-1} = \begin{pmatrix} x_{j_1}, & \cdots, & x_{j_m} \\ k_1^{-1} x_1, & \cdots, & k_m^{-1} x_m \end{pmatrix}. \tag{3}$$

By this definition of the multiplication, the set of all monomial substitutions on m variables is seen to form a group $\Sigma_m(H)$. This group we shall call the *complete monomial group* of degree m, or somewhat more simply, the *symmetry* of degree m of H. If H is a finite group of order n_H then one finds that $\Sigma_m(H)$ has the order $m!(n_H)^m$. The ordinary symmetric group can be considered the symmetry of the unit group.

A *permutation* in Σ_m is a monomial substitution of the form

$$\pi = \begin{pmatrix} x_1, & \cdots, & x_m \\ x_{i_1}, & \cdots, & x_{i_m} \end{pmatrix} = \begin{pmatrix} 1, & \cdots, & m \\ i_1, & \cdots, & i_m \end{pmatrix}. \tag{4}$$

The permutations form a subgroup S_m of Σ_m and S_m is obviously isomorphic to the ordinary symmetric group on m letters.

A monomial substitution which only multiplies each variable by a factor in H

$$\mu = \begin{pmatrix} x_1, & \cdots, & x_m \\ r_1 x_1, & \cdots, & r_m x_m \end{pmatrix} = [r_1, \cdots, r_m] \tag{5}$$

will be called a *multiplication*. The multiplications μ form a subgroup $V_m(H)$ of $\Sigma_m(H)$ which we shall call the *basis group*. It is easily seen that the basis group is a normal subgroup of $\Sigma_m(H)$. Furthermore it is the direct product of m groups isomorphic to H

$$V_m = H_1^* \times \cdots \times H_m^* \tag{6}$$

where H_i^* consists of the multiplications

$$\eta^{(i)} = [1, \cdots, 1, r_i, 1, \cdots, 1] \tag{7}$$

where r_i runs through H.

The special multiplications

$$\mu = [a, a, \cdots, a] = [a]$$

where all factors are equal, will be called *scalars*. The scalars are the only elements which commute with all permutations. The *center* of the symmetry consists of all scalars

$$\sigma = [c, c, \cdots, c]$$

where c belongs to the center of H. This shows that the centers of H and $\Sigma_m(H)$ are isomorphic groups.

Any monomial substitution can be written uniquely as the product of a permutation and a multiplication. If ρ, π and μ are given by (1), (4) and (5), respectively, one finds $\rho = \pi\mu$. This shows that one has

$$\Sigma_m = S_m \cup V_m, \qquad S_m \cap V_m = E. \tag{8}$$

2. **Auxiliary results.** Before we proceed to a more detailed study of the symmetries, it is necessary to derive a few auxiliary theorems which also have some independent interest.

We first prove the following theorem:

THEOREM 1. *Let $d_1, d_2, \cdots, d_n$ be n elements of a group H all belonging to the same class. Then there exist elements c_i in H such that*

$$d_1 = c_1 \cdots c_n,\ d_2 = c_2 \cdots c_n c_1, \cdots, d_n = c_n c_1 \cdots c_{n-1}. \tag{9}$$

Proof. According to the assumption we may write

$$d_2 = c_1^{-1} d_1 c_1,\ d_3 = c_2^{-1} d_2 c_2, \cdots, d_n = c_{n-1}^{-1} d_{n-1} c_{n-1}.$$

Then is is only necessary to determine the element c_n such that $d_1 = c_1 \cdots c_{n-1}c_n$ and one finds that all other relations (9) are satisfied.

For $n=2$ the theorem expresses the fact that if d_1 and d_2 belong to the same class then there exist elements c_1 and c_2 in H such that $d_1 = c_1c_2$, $d_2 = c_2c_1$.

Next we prove:

THEOREM 2. *Let the products $a_1a_2 \cdots a_n$, $b_1b_2 \cdots b_n$ belong to the same class in a group H. Then there exist elements $q_1, \cdots, q_n$ in H such that*

$$b_1 = q_1^{-1} a_1 q_2,\ b_2 = q_2^{-1} a_2 q_3, \cdots, b_n = q_n^{-1} a_n q_1. \tag{10}$$

Proof. If one writes $b_1 \cdots b_n = q_1^{-1}(a_1 \cdots a_n)q_1$ then one finds that the relations (10) can be satisfied by putting

$$q_i = a_{i-1}^{-1} \cdots a_1^{-1} q_1 b_1 \cdots b_{i-1}.$$

3. **Cycles and transformations.** In the theory of monomial groups as well as in the theory of substitution groups it is advantageous to introduce cycles of monomial substitutions, i.e., substitutions having the form

$$\gamma = \begin{pmatrix} x_1, & x_2, & \cdots, & x_n \\ c_1x_2, & c_2x_3, & \cdots, & c_nx_1 \end{pmatrix} = (c_1x_2, c_2x_3, \cdots, c_{n-1}x_n, c_nx_1). \tag{11}$$

As in the theory of permutation groups one shows:

THEOREM 3. *A monomial substitution can be written uniquely as the product of commutative cycles without common variables.*

One finds that the nth power of a cycle (11) of length n is a multiplication

$$\gamma^n = \begin{pmatrix} x_1, & \cdots, & x_n \\ \Delta_1 x_1, & \cdots, & \Delta_n x_n \end{pmatrix} = [\Delta_1, \Delta_2, \cdots, \Delta_n]. \tag{12}$$

Here the factors Δ_i are determined by

$$\Delta_1 = c_1 c_2 \cdots c_n, \Delta_2 = c_2 \cdots c_n c_1, \cdots, \Delta_n = c_n c_1 \cdots c_{n-1}. \tag{13}$$

These elements in H will be called the *determinants* of γ. Since one can write the relations (13) in the form

$$\Delta_2 = c_1^{-1} \Delta_1 c_1, \cdots, \Delta_n = c_{n-1}^{-1} \Delta_{n-1} c_{n-1}, \Delta_1 = c_n^{-1} \Delta_n c_n, \tag{14}$$

one concludes that with each cycle (11) there is associated a unique *determinant class* in H.

From (12) one concludes:

THEOREM 4. *The order of a cycle γ of length n is nt where t is the order of its determinant class. The order of any monomial substitution is the least common multiple of the orders of its cycles.*

We shall next prove the following fact.

THEOREM 5. *Any elements $d_1, \cdots, d_n$ in the same class in H are the determinants of some monomial substitution of length n.*

Proof. This theorem is a restatement of Theorem 1.

We shall now turn to the transformation of one monomial substitution by another. Obviously it is sufficient to study what happens to a single cycle (11) by transformation. From (2) and (3) one obtains

$$\kappa\gamma\kappa^{-1} = \begin{pmatrix} x_{j_1}, & x_{j_2}, & \cdots, & x_{j_n} \\ k_1^{-1} c_1 k_2 x_{j_2}, & k_2^{-1} c_2 k_3 x_{j_3}, & \cdots, & k_n^{-1} c_n k_1 x_{j_1} \end{pmatrix}. \tag{15}$$

This shows that every cycle goes into a cycle of the same length. Furthermore the determinants of the transformed cycle (15) are similar to those of γ, namely

$$\Delta_1' = k_1^{-1} \Delta_1 k_1, \cdots, \Delta_n' = k_n^{-1} \Delta_n k_n. \tag{16}$$

This proves that the determinant class of a cycle is invariant under transformation.

THEOREM 6. *The necessary and sufficient condition for two monomial cycles to be similar is that they shall have the same length and the same determinant class.*

Proof. It has already been shown that the condition is necessary and the sufficiency follows from (15) and Theorem 2.

From Theorem 6 one draws the more general conclusion that two mono-

mial substitutions are similar if and only if the cycles in their cyclic decompositions may be made to correspond in such a manner that corresponding cycles have the same length and determinant class.

We show next:

THEOREM 7. *Any cycle of length n may be transformed to the normal form*

$$\gamma = \begin{pmatrix} x_{i_1}, & \cdots, & x_{i_{n-1}}, & x_{i_n} \\ x_{i_2}, & \cdots, & x_{i_n}, & ax_{i_1} \end{pmatrix} = (x_{i_2}, \cdots, x_{i_n}, ax_{i_1}) \tag{17}$$

where a is any element in the determinant class of γ. Any monomial substitution ρ is similar to a product of cycles without common variables $\rho = \gamma_1 \cdots \gamma_r$ where each cycle is in normal form.

Proof. The determinant class of the cycle (17) is the same as the class of a, and hence a cycle may be transformed into the form (17) by Theorem 6. One sees also directly from (15) that a cycle may be transformed in such a way that the $n-1$ first factors become the unit element. Since the transformation of γ into the form (17) may be performed by means of a substitution involving only the same variables, all cycles in ρ may be transformed into normal form simultaneously.

4. **Centralizers.** We shall use the preceding results to determine the centralizer of an arbitrary element ρ in the symmetry. Since similar elements have similar centralizers one can assume that ρ is in the normal form of Theorem 7.

The centralizer shall be determined by the investigation of a few special cases. We assume first that $\rho = \gamma_a$ is a single cycle

$$\gamma_a = (ax_1, x_2, \cdots, x_n). \tag{18}$$

When γ_a is transformed by κ in (2) one finds as in (15)

$$\kappa\gamma_a\kappa^{-1} = (k_n^{-1}ak_1x_{j_1}, k_1^{-1}k_2x_{j_2}, \cdots, k_{n-1}^{-1}k_nx_{j_n}). \tag{19}$$

Since this cycle will be identical with (18) when κ belongs to the centralizer one sees first that κ must have the form

$$\kappa = \begin{pmatrix} x_1, & x_2, & \cdots, & x_{n-j+1}, & x_{n-j+2}, & \cdots, & x_n \\ k_1x_j, & k_2x_{j+1}, & \cdots, & k_{n-j+1}x_n, & k_{n-j+2}x_1, & \cdots, & k_nx_{j-1} \end{pmatrix}$$

while the substitution on the remaining $m-n$ variables is immaterial. For this reason we consider only the centralizer C_γ of γ in $\Sigma_n(H)$. The values of the factors k_i in κ can be obtained by comparison of (18) and (19) and one finds without difficulty that an arbitrary element κ in the centralizer C_γ has the form

$$\kappa = \begin{pmatrix} x_1, & \cdots, & x_{n-j+1}, & x_{n-j+2}, & \cdots, & x_n \\ cx_j, & \cdots, & cx_n, & cax_1, & \cdots, & cax_{j-1} \end{pmatrix}$$

where c is an arbitrary element in the centralizer of a in H. Obviously the powers of the cycle γ_a must belong to C_γ and one sees directly that

$$\kappa = [c]\gamma_a^j = \gamma_a^j[c]$$

where $[c]$ is the multiplication in which every variable is multiplied by c. This shows that the centralizer C_γ is isomorphic to a cyclic extension of degree n of a group isomorphic to the centralizer of a in H. If this last centralizer is finite and of order r_a then one sees that the order of C_γ is nr_a.

Next let us determine the centralizer C_λ of a product of cycles of the same length n and the same determinant class a,

$$\lambda = \gamma^{(1)} \cdots \gamma^{(k)}, \tag{20}$$

in the symmetry of degree corresponding to the variables involved. Obviously any permutation of the cycles $\gamma^{(i)}$ in (20) among themselves must belong to the centralizer C_λ, and an element of the centralizer leaving all cycles in ρ fixed by transformation must have the form

$$c_\lambda = c^{(1)} \cdots c^{(k)}$$

where $c^{(i)}$ belongs to the centralizer of $\gamma^{(i)}$ in the symmetry on its variables. From these results one concludes that C_λ is isomorphic to the symmetry $\Sigma_k(C_\gamma)$, where C_γ is the centralizer of a single cycle γ.

Since any element ρ in $\Sigma_m(H)$ can be written as a product of cycles γ when it is transformed to normal form, one finally obtains:

THEOREM 8. *Let an element in $\Sigma_m(H)$ be transformed into its normal form*

$$\rho = \lambda_1 \cdots \lambda_t, \qquad \lambda_i = \gamma_1^{(i)} \cdots \gamma_{k_i}^{(i)}$$

where for a fixed i the $\gamma_j^{(i)}$ are the normalized cycles of the same length n_i and the same determinant class a_i. Then the centralizer C_ρ of ρ in $\Sigma_m(H)$ is isomorphic to the direct product of symmetries

$$C_\rho \simeq \Sigma_{k_1}(C_{a_1}) \times \Sigma_{k_2}(C_{a_2}) \times \cdots$$

where C_{a_i} is the centralizer of a single cycle $\gamma_j^{(i)}$ in $\Sigma_{n_i}(H)$. The group C_{a_i} consists of all elements of the form $\kappa = [c_i]\gamma_1^{(i)^j}$, where the element c_i belongs to the normalizer of a_i in H.

Again if r_i is the order of the centralizer of a_i in H one finds that the order of C_ρ is

$$N = \prod_i k_i!(n_i r_i)^{k_i}.$$

5. **Splitting of the symmetry.** A group G containing a normal subgroup N is said to *split* over N if there exists a subgroup M such that

$$G = M \cup N, \qquad M \cap N = E.$$

In this representation the group M can be replaced by any of its conjugates and the relations will still hold. There may, however, exist other groups M_1 such that

$$G = M_1 \cup N, \qquad M_1 \cap N = E$$

and such that M_1 is not a conjugate of M. This leads to a division of all representative groups M into classes, each consisting of conjugate groups. When there is only one class, hence when all M are conjugate, we say that G *splits regularly* over N.

We have already seen in (8) that for the symmetry one has

$$\Sigma_m = S_m \cup V_m, \qquad S_m \cap V_m = E \tag{21}$$

and hence Σ_m splits over V_m. We shall now consider the problem of finding all groups T such that

$$\Sigma_m = T \cup V_m, \qquad T \cap V_m = E. \tag{22}$$

One sees immediately that T is a group isomorphic to S_m and the isomorphism is such that each element in T is obtained from the corresponding element in S_m by multiplication with an element in the basis group V_m.

The group S_m contains the transpositions $(1, i)$. This implies that T contains $m-1$ substitutions of the form

$$\lambda_i = [a_{1,i}, \cdots, a_{m,i}](x_1, x_i). \tag{23}$$

Let us transform all λ_i by the multiplication

$$\kappa = [k_1, \cdots, k_m].$$

One finds that the group $\kappa T \kappa^{-1}$ contains all the substitutions

$$\lambda_i' = [\overset{-1}{k_i} a_{1,i} k_1, \cdots, \overset{-1}{k_1} a_{i,i} k_i, \cdots, \overset{-1}{k_j} a_{j,i} k_j, \cdots](x_1, x_i)$$

where j is an index different from 1 and i. This shows that it is possible to choose κ in such a manner that $a_{1,i} = 1$ for every i, and in the following we shall make this assumption.

Next one finds

$$\lambda_i^2 = [a_{i,i}, a_{2,i}^2, \cdots, a_{i,i}, \cdots, a_{m,i}^2]$$

and since it is a multiplication contained in T one must have $\lambda_i^2 = E$; hence

$$a_{i,i} = a_{1,i} = 1, \qquad a_{i,j}^2 = 1. \tag{24}$$

In the special case $m=2$ it follows already from these conditions that $\lambda_2=(1, 2)$ and we have shown:

THEOREM 9. *For $m=2$ the symmetry splits regularly over its basis group.*

In the following we can assume $m\geqq 3$. Let us form the product

$$\lambda_i\lambda_j = \begin{pmatrix} x_1, & x_j, & x_i, & \cdots, & x_k, & \cdots \\ a_{j,i}x_j, & x_i, & a_{i,j}x_1, & \cdots, & a_{k,j}a_{k,i}x_k, & \cdots \end{pmatrix}$$

where k is any index different from 1, i and j. Since the third power of this substitution must also be the unit substitution one obtains the conditions

$$a_{j,i}a_{i,j} = 1. \qquad (a_{k,j}a_{k,i})^2 = 1, \qquad k \neq i,\ k \neq j.$$

When the first of these conditions is combined with (24) one obtains

$$a_{i,j} = a_{j,i}. \tag{25}$$

The results expressed in (23) and (25) are already sufficient to give a complete solution for the case $m=3$. One finds then

$$\lambda_2 = \begin{pmatrix} x_1, & x_2, & x_3 \\ x_2, & x_1, & ax_3 \end{pmatrix}, \qquad \lambda_3 = \begin{pmatrix} x_1, & x_2, & x_3 \\ x_3, & ax_2, & x_1 \end{pmatrix}. \tag{26}$$

These two elements generate a group isomorphic to S_3 consisting of the substitutions (26), the unit element and the following three substitutions:

$$\begin{gathered} \mu = \begin{pmatrix} x_1, & x_2, & x_3 \\ ax_3, & ax_1, & x_2 \end{pmatrix}, \qquad \mu^2 = \begin{pmatrix} x_1, & x_2, & x_3 \\ ax_2, & x_3, & ax_1 \end{pmatrix}, \\ \nu = \begin{pmatrix} x_1, & x_2, & x_3 \\ ax_1, & ax_3, & ax_2 \end{pmatrix}. \end{gathered} \tag{27}$$

From now on we can even assume $m\geqq 4$. We evaluate the product

$$\lambda_j\lambda_i\lambda_j = \begin{pmatrix} x_1, & x_j, & x_i, & \cdots, & x_k, & \cdots \\ a_{i,j}x_1, & a_{i,j}x_i, & a_{i,j}x_j, & \cdots, & a_{k,j}a_{k,i}a_{k,j}x_k, & \cdots \end{pmatrix} \tag{28}$$

where k denotes any index different from $1, i, j$. In the isomorphism between S_m and T the substitution λ_i corresponds to the transposition $(1, i)$ in S_m, and hence $\lambda_j\lambda_i\lambda_j$ must correspond to (i, j).

Next we utilize the fact that in S_m

$$(1, h)(i, j) = (i, j)(1, h)$$

for four different indices $1, h, i, j$. This implies for the corresponding elements in T

$$\lambda_h \cdot \lambda_j\lambda_i\lambda_j = \lambda_j\lambda_i\lambda_j \cdot \lambda_h.$$

When these products are compared one finds that they imply the identities

$$a_{i,j} = a_{h,j}a_{h,i}a_{h,j}, \qquad a_{i,j}a_{k,h} = a_{k,h}a_{i,j}.$$

When the first of these relations is applied to (28) one finds the simpler expression

$$\lambda_j\lambda_i\lambda_j = [a_{i,j}]\begin{pmatrix} x_i, & x_j \\ x_j, & x_i \end{pmatrix}. \tag{29}$$

This important relation shows that if $i\neq 1$, $j\neq 1$, then the element in T corresponding to (i, j) in S_m is obtained simply through the multiplication of (i, j) by a scalar $[a_{i,j}]$. But the transpositions (i, j) generate the group of all permutations π_1 leaving the element 1 fixed; hence one has for any such permutation

$$\pi_1 \to \pi_1[a_{\pi_1}].$$

Since the scalars commute with all permutations this shows that the a_{π_1} form a subgroup of H which is homomorphic to the symmetric group on $m-1$ letters.

This leads to the following theorem:

THEOREM 10. *The symmetry $\Sigma_m(H)$ splits over its basis group $\Sigma_m = T\cup V_m$, $T\cap V_m = E$. Any group T in this decomposition is the conjugate of some group T_0 obtained by the following construction. Let A be any subgroup of H homomorphic to the symmetric group on $m-1$ letters, and let $\pi_1 \to a_{\pi_1}$ indicate the homomorphism between A and the permutations π_1 in the symmetric group $S_{m-1}^{(1)}$ on the $m-1$ letters $2, \cdots, m$. In particular we shall write*

$$(i, j) \to a_{i,j}, \qquad (i, i) \to 1 = a_{i,i}.$$

Then the elements of T_0 are obtained from the permutations of the symmetric group S_m by the isomorphism

$$\begin{aligned} \pi_1 &\to \pi_1[a_{\pi_1}] \quad \text{for} \quad \pi_1 \text{ in } S_{m-1}^{(1)}, \\ (1, i) &\to \lambda_i = [1, a_{2,i}, \cdots, a_{m,i}](1, i). \end{aligned} \tag{30}$$

It has already been shown that any group T after suitable transformation must have the form indicated by the correspondence (30). It remains to show that the set of substitutions defined by the correspondence (30) actually forms a group isomorphic to S_m.

We observe first that any permutation in S_m can be written uniquely in the form

$$\pi = (1, i)\pi_1 \tag{31}$$

where π_1 leaves 1 fixed. Thus (30) defines a unique substitution T_π corresponding to each permutation π, namely,

$$T_\pi = T_{(1,i)} \cdot T_{\pi_1} \tag{32}$$

or

$$T_\pi = [1, a_{2,i}, \cdots, a_{m,i}][a_{\pi_1}](1, i)\pi_1$$

and therefore

$$T_\pi = [a_{\pi_1}, a_{2,i}a_{\pi_1}, \cdots, a_{m,i}a_{\pi_1}]\pi. \tag{33}$$

Our problem is to prove that this correspondence is an isomorphism. Let $\rho = (1, j)\rho_1$ be another permutation in the normal form (31) where ρ_1 leaves 1 fixed. We shall then have to prove

$$T_\pi \cdot T_\rho = T_{\pi\rho}. \tag{34}$$

From the definition (33) of T_π it follows that

$$T_{\pi\rho_1} = T_\pi \cdot T_{\rho_1}$$

for any π and any ρ_1 leaving 1 fixed. When this is applied to (34) one sees that it is no limitation to assume simply $\rho = (1, j)$.

We shall have to prove, therefore, that

$$T_{\pi(1,j)} = T_\pi \cdot T_{(1,j)} = T_{(1,i)} \cdot T_{\pi_1} \cdot T_{(1,j)}$$

and this follows if one can prove the two simpler relations

$$T_{(1,i)} \cdot T_\rho = T_{(1,i)\rho} \tag{35}$$

for any permutation ρ, and

$$T_{\pi_1\rho} = T_{\pi_1} \cdot T_\rho \tag{36}$$

for any π_1 leaving 1 fixed.

Let us first turn our attention to the relation (35). By the preceding remark it is sufficient to prove

$$T_{(1,i)} \cdot T_{(1,j)} = T_{(1,i)(1,j)}. \tag{37}$$

For $i = j$ this relation follows simply from $a_{i,j}^2 = 1$. We assume therefore $i \neq j$ and find

$$T_{(1,i)} \cdot T_{(1,j)} = [1, \cdots, a_{t,i}, \cdots](1, i)[1, \cdots, a_{t,j}, \cdots](1, j)$$

or

$$T_{(1,i)} \cdot T_{(1,j)} = \begin{pmatrix} x_1, & x_i, \cdots, x_t, & \cdots \\ a_{i,j}x_1, & x_i, \cdots, a_{t,j}a_{t,i}x_t, & \cdots \end{pmatrix} (1, i)(1, j)$$

where t is any index different from 1 and i. On the other hand one obtains

$$T_{(1,i)(1,j)} = T_{(1,j,i)} = T_{(1,j)(i,j)} = [1, \cdots, a_{t,j}, \cdots][a_{i,j}](1, i)(1, j)$$

or

$$T_{(1,i)(1,j)} = \begin{pmatrix} x_1, & x_i, \cdots, x_t, & \cdots \\ a_{i,j}x_1, & x_i, \cdots, a_{i,j}a_{t,j}x_t, & \cdots \end{pmatrix}(1, i)(1, j).$$

Since $a_{i,j}=a_{j,i}$, $a_{t,j}a_{t,i}=a_{i,j}a_{t,j}$ the relation (37) follows.

To prove the relation (36) let us observe that π can be written as the product of transpositions not involving 1. If (36) is proved for any transposition π_1 it will therefore follow in general. By the preceding remarks we can also assume $\rho=(1, j)$, and hence the proof of (36) reduces to

$$T_{(k,l)} \cdot T_{(1,j)} = T_{(k,l)(1,j)}. \tag{38}$$

Here we may assume $k \neq l$ and one needs only to study the two cases $j \neq l$ and $j=l$ since k and l occur symmetrically. A simple calculation which we shall not reproduce here shows that (38) actually holds in both cases.

This concludes the proof of Theorem 10. It should be observed that this theorem is of interest since it gives complete information about a splitting extension in a general and rather involved case.

To conclude let us prove:

THEOREM 11. *The necessary and sufficient condition for the symmetry $\Sigma_m(H)$ to split regularly over its basis group $V_m(H)$ is that H contain no subgroup different from E homomorphic to the symmetric group on $m-1$ letters.*

Proof. If H contains no such group, all the factors $a_{i,j}$ in Theorem 10 are the unit element and the group constructed is the symmetric group S_m.

Conversely if Σ_m splits regularly the group T can be transformed into S_m. Since the transformation of a substitution in Σ_m by a permutation only permutes its factors it follows that this transformation is a multiplication,

$$\kappa = [k_1, \cdots, k_m].$$

But when κ transforms

$$\lambda_i = [1, a_{2,i}, \cdots, a_{m,i}](1, i)$$

into a permutation one finds that all k_i are equal and $a_{i,j}=1$.

CHAPTER II. NORMAL SUBGROUPS OF THE SYMMETRY

1. **Permutation-invariant subgroups.** The main problem to be considered in the following is the determination of all normal subgroups of the symmetry $\Sigma_m(H)$. Before this problem can be solved it is however necessary to solve a preliminary problem which also has some independent interest.

We shall say that a subgroup of the symmetry is *permutation invariant* if it is transformed into itself by all permutations π of the symmetric group S_m. The first problem to be considered is then *the determination of all permutation invariant subgroups of the basis group V_m.*

Let $\overline{N}$ be a permutation invariant subgroup of V_m,

$$\kappa = [k_1, \cdots, k_m] \tag{1}$$

one of its elements and

$$\pi = \begin{pmatrix} 1, & \cdots, & m \\ i_1, & \cdots, & i_m \end{pmatrix} \tag{2}$$

an arbitrary permutation. Since $\overline{N}$ is permutation invariant it must also contain all multiplications

$$\kappa_\mu = \pi^{-1}\kappa\pi = [k_{i_1}, \cdots, k_{i_m}] \tag{3}$$

and conversely if $\overline{N}$ contains all multiplications (3) when it contains κ, $\overline{N}$ is permutation invariant. One sees easily that if $\overline{N}$ and $\overline{M}$ are permutation invariant subgroups of V_m then $\overline{N}\cup\overline{M}$ and $\overline{N}\cap\overline{M}$ have the same property, and hence the permutation invariant subgroups of V_m form a structure.

Let us mention three simple types of permutation invariant subgroups of V_m.

1. Subgroups $\overline{N}$ where the k_i in (1) run through all elements in a certain subgroup K of H independently. Then $\overline{N}$ is isomorphic to the direct product of m groups, each isomorphic to K.

2. The k_i run through K in such a manner that $k_1 = k_2 = \cdots = k_m$. In this case $\overline{N}$ is isomorphic to K.

3. K is Abelian and the k_i are so restricted that $k_1k_2 \cdots k_m = e$. In this case $\overline{N}$ is isomorphic to the direct product of $m-1$ factors, each isomorphic to K.

In the following we shall show that all permutation invariant subgroups of V_m may be obtained essentially through a combination of these three types.

The proposed problem may also be formulated in a slightly different manner. We have already seen that the basis group V_m is isomorphic to the direct product

$$V_m \cong H^{(1)} \times \cdots \times H^{(m)}$$

where each group $H^{(i)}$ is isomorphic to H. Now conversely let m such groups $H^{(i)}$ be given and let T_i denote the isomorphism which takes H into $H^{(i)}$. Then the direct product $P = H^{(1)} \times \cdots \times H^{(m)}$ consists of all elements of the form

$$p = h_1^{T_1} h_2^{T_2} \cdots h_m^{T_m}. \tag{4}$$

Our problem is then equivalent to the determination of all those subgroups of P which have the property that when (4) is contained in the subgroup, then it also contains all the products $p' = h_{i_1}^{T_1} h_{i_2}^{T_2} \cdots h_{i_m}^{T_m}$ obtained by an arbitrary permutation of the indices.

2. **Permutation invariant subgroups for** $m=2$. From now on let $\overline{N}$ denote a fixed permutation invariant subgroup of the basis group V_m and let κ in (1) be one of its elements. All elements k_i of H which occur as the ith factor in any such multiplication will form a subgroup K_i of H. But since $\overline{N}$ is permutation invariant this group must be the same, say K, for all indices i. One sees then that in order to determine $\overline{N}$ it is necessary to know K and the rule which determines which elements k_i in K may be combined to give an element (1) of $\overline{N}$.

There exist certain subgroups of the permutation invariant subgroup $\overline{N}$ of V_m which are of particular importance. Those elements of $\overline{N}$ which have the form

$$\sigma_i = [s_1, \cdots, s_i, 1, \cdots, 1] \tag{5}$$

form a subgroup $\overline{S}_i$ of $\overline{N}$. When σ_i is transformed by κ in (1) one finds

$$\kappa\sigma_i\kappa^{-1} = [k_1^{-1}s_1k_1, \cdots, k_i^{-1}s_ik_i, 1, \cdots, 1].$$

This shows that $\overline{S}_i$ is normal in $\overline{N}$. Furthermore those s_j in (5) which occur as factors at the jth place also form a subgroup S_i of H independent of j and S_i is normal in K.

When the various permutations π in (2) are applied to $\overline{S}_i$ one obtains $C_{m,i}$ other subgroups $\overline{S}_i^{(r)}$ leaving i variables unchanged. Among these groups we shall consider first the m groups $\overline{S}_1^{(j)}$ ($j=1, 2, \cdots, m$) consisting of the elements of the form

$$[1, \cdots, 1, s_j, 1, \cdots, 1]. \tag{6}$$

We have already seen that the corresponding s_j form a normal subgroup S_1 of K. But then the direct product

$$P_1 = \overline{S}_1^{(1)} \times \cdots \times \overline{S}_1^{(m)}$$

is a normal subgroup of $\overline{N}$. Furthermore P_1 is permutation invariant of the first type indicated in §1 and obviously it is the largest such group contained in $\overline{N}$. This means that if κ in (1) is an arbitrary element of $\overline{N}$ then any k_i may be multiplied by an arbitrary s_i in S_1 and still κ remains in $\overline{N}$. This shows that the eventual conditions on the relations between the various k_i in an element (1) in $\overline{N}$ can only be conditions mod S_1. It is therefore no limitation if we consider the quotient group K/S_1 in the following and therefore assume $S_1=1$; hence $s_j=1$ in (6).

The case $m=2$ presents a certain exception and we shall therefore solve our problem first in this case. If $\overline{N}$ contains two elements

$$\alpha = [a_1, a_2], \qquad \beta = [a_1, b_2]$$

with the same first component a_1 then

$$\alpha\beta^{-1} = [1, b_2^{-1}a_2] = [1, 1]$$

and hence $a_2 = b_2$. This shows that a_1 determines a_2 uniquely and conversely. Furthermore the correspondence between a_1 and a_2 is an automorphism T of K: $a_2 = a_1^T$. But since $\overline{N}$ is permutation invariant, a_1 and a_2 may be interchanged and one finds $a_1 = a_2^T$, and T is an automorphism of order 2. Conversely one finds that the multiplications $[a_1, a_1^T]$ do form a permutation invariant subgroup of V_2 if T is an automorphism of K of order 2. We have therefore shown:

THEOREM 1. *All permutation invariant subgroups $\overline{N}$ of $V_2(H)$ can be obtained by the following construction. A subgroup K of H is chosen. S_1 denotes an arbitrary normal subgroup of K while T is a fixed automorphism of order 2 of the quotient group K/S_1. Then $\overline{N}$ consists of all elements*

$$\alpha = [a_1, a_2]$$

where a_1 and a_2 run through K in such a manner that

$$a_2 \equiv a_1^T \pmod{S_1}.$$

3. **Permutation invariant subgroups of the basis group in the general case.** We shall now turn to the determination of the permutation invariant subgroups $\overline{N}$ of $V_m(H)$ in the general case where $m \geqq 3$. Let us recall also that we have assumed $S_1 = 1$. We shall first study in some detail the properties of the group $\overline{S}_2$ which was defined in §2 as the set of all elements of $\overline{N}$ of the form $\alpha_2 = [s_1, s_2, 1, \cdots, 1]$. Since $\overline{N}$ is permutation invariant, $\overline{S}_2$ must be unchanged when s_1 and s_2 are interchanged. Theorem 1 may therefore be applied and one concludes that

$$\alpha_2 = [s_1, s_1^T, 1, \cdots, 1] \tag{7}$$

where T is some automorphism of order 2 of the group S_2.

When α_2 in (7) is transformed by an arbitrary permutation one finds that for any element

$$[1, \cdots, 1, s_i, 1, \cdots, 1, s_j, 1, \cdots, 1]$$

in $\overline{N}$ one has $s_j = s_i^T$. But when $\overline{N}$ contains (7) it also contains

$$\beta_2 = [s_1, 1, s_1^T, 1, \cdots, 1]$$

and hence the quotient

$$\alpha_2\beta_2^{-1} = [1, s_1^T, (s_1^{-1})^T, 1, \cdots, 1],$$

and one finds

$$s_1^T = s_1^{-1}.$$

But if a group has an automorphism changing each element into its inverse then $(s_1s_2)^T=(s_1s_2)^{-1}=s_1^Ts_2^T=s_1^{-1}s_2^{-1}$ or $s_2s_1=s_1s_2$ and the group is Abelian. We have therefore obtained the following result:

The group $\overline{S}_2$ consists of all elements

$$\alpha_2 = [s_1, \overset{-1}{s_1}, 1, \cdots, 1] \tag{8}$$

where s_1 runs through a normal Abelian subgroup of K.

When the elements s_1 and s_1^{-1} in (8) are permuted into all possible positions, the corresponding substitutions generate a normal subgroup $\overline{R}$ of $\overline{N}$ which is also permutation invariant. From the construction of this group it follows that if $\rho=[r_1, \cdots, r_m]$ is one of its elements, then

$$r_1r_2 \cdots r_m = 1, \tag{9}$$

and hence $\overline{R}$ is a group of the third type indicated in §1. From the definition of $\overline{R}$ it follows that it contains all substitutions

$$\rho_1 = [p_1, \overset{-1}{p_1}, 1, \cdots, 1], \qquad \rho_2 = [1, p_2, \overset{-1}{p_2}, 1, \cdots, 1],$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots ,$$
$$\rho_{m-1} = [1, \cdots, 1, p_{m-1}, \overset{-1}{p_{m-1}}],$$

where the p_i are arbitrary elements in S_2. But then $\overline{R}$ contains the product

$$\gamma = \rho_1\rho_2 \cdots \rho_{m-1}$$

or

$$\gamma = [p_1, p_2\overset{-1}{p_1}, \cdots, p_{m-1}\overset{-1}{p_{m-2}}, \overset{-1}{p_{m-1}}] \tag{10}$$

and one sees easily that every element in $\overline{R}$ can be expressed in this form. By a suitable choice of the elements p_i in (10) one can make the first $m-1$ factors arbitrary in S_2 while the last is uniquely determined by the condition (9). The following has been shown therefore:

The group $\overline{R}$ consists of all elements of the form

$$\rho = [r_1, \cdots, r_{m-1}, (r_1 \cdots r_{m-1})^{-1}] \tag{11}$$

where the r_i run through the Abelian group S_2 independently. The group $\overline{R}$ is isomorphic to the direct product of $m-1$ factors each isomorphic to S_2.

We can now turn to the final step in the determination of the permutation invariant subgroups $\overline{N}$ of V_m. Let κ in (1) be an arbitrary element of $\overline{N}$ and let

$$\lambda = [k_2, k_1, k_3, \cdots, k_m]$$

be the element obtained by interchanging k_1 and k_2. Then one finds

$$\kappa\overset{-1}{\lambda} = [\overset{-1}{k_2}k_1, \overset{-1}{k_1}k_2, 1, \cdots, 1]$$

showing that the quotient belongs to $\overline{S}_2$, and hence $k_2^{-1}k_1$ is an element of S_2. Since the same result can be obtained for any $k_i^{-1}k_1$ one concludes that the elements κ of $\overline{N}$ must have the form

$$\kappa = [kr_1, kr_2, \cdots, kr_m] \tag{12}$$

where all the r_i are elements of S_2 and k runs through K.

It remains to determine the relations between the r_i in (12) which will insure that these elements form a permutation invariant group. Since $\overline{R}$ contains (11) one finds that $\overline{N}$ must contain the element

$$\rho^{-1}\kappa = (k, k, \cdots, k, k'). \tag{13}$$

All elements of this form in $\overline{N}$ must form a group and the element k' is uniquely determined by k. This correspondence must be a homomorphism and one can write therefore $k' = k^T$ where T is an endomorphism of K. When applied to (11) this gives the final form

$$\kappa = [kr_1, \cdots, kr_{m-1}, k^T(r_1 \cdots r_{m-1})^{-1}] \tag{14}$$

for the elements belonging to $\overline{N}$.

In (13) one has $k' = kr$, where r belongs to S_2, and hence the endomorphism T of K must transform any element k into an element obtained by multiplication of k by an element in S_2. When one considers the special element $[r, \cdots . r, r^{-(m-1)}]$ in $\overline{R}$ it follows that

$$r^T = r^{-(m-1)}$$

for any element r in S_2. This relation shows that in (12) the last factor is uniquely determined by the $m-1$ first ones.

Conversely one verifies that under the stated conditions on T and under the assumption that S_2 is an Abelian group, the elements (14) form a group. Furthermore if two factors in (14) are permuted the quotient between the corresponding substitutions also belongs to the group, and hence it is permutation invariant. We have therefore obtained the principal result:

THEOREM 2. *Let $m \geqq 3$. Then all permutation invariant subgroups $\overline{N}$ of $V_m(H)$ may be constructed by the following process. A subgroup K of H is chosen. In K two normal subgroups $S_2 \supset S_1$ are selected such that the quotient group S_2/S_1 is Abelian. Then $\overline{N}$ consists of the elements*

$$\kappa = [k_1, \cdots, k_m]$$

where the k_i run through K subject to the conditions

$$\left.\begin{aligned} k_i &\equiv ks_i \ (i = 1, 2, \cdots, m-1), \\ k_m &\equiv k^T \ (s_1 \cdots s_{m-1})^{-1}, \end{aligned}\right\} \pmod{S_1},$$

where the s_i are arbitrary elements in S_2. Furthermore T is an endomorphism of

K/S_1 multiplying each element of K/S_1 by an element in S_2/S_1 and in particular

$$s^T \equiv s^{-(m-1)} \pmod{S_1}$$

for any element s in S_2.

Here we have solved the problem of finding all subgroups of $V_m(H)$ which are invariant under all permutations of the symmetric group. A much more difficult problem is the determination of those subgroups of $V_m(H)$ which are invariant under the permutations of a fixed subgroup of the symmetric group.

4. **Normal subgroups of the symmetry contained in the basis group.** We shall now turn to the determination of the normal subgroups of the symmetry and we consider first those normal subgroups of $\Sigma_m(H)$ which are contained in the basis group. Since we have observed in the preceding that every substitution in Σ_m is the product of a permutation and a multiplication, we shall have to determine those normal subgroups of V_m which are permutation invariant.

We use the same notations as before. If (1) is an element of a normal subgroup $\overline{N}$ of V_m the factors k_i must run through a normal subgroup K of H. Furthermore the subgroups $\overline{S}_i$ defined by the elements (5) must also be normal in V_m and the corresponding factors s_i run through a normal subgroup S_i of H.

The case $m=2$ must again be considered separately. When $\overline{N}$ contains the element $\kappa=[k_1, k_2]$ it must also contain the element $\kappa'=[hk_1h^{-1}, k_2]$ for an arbitrary h in H. But in Theorem 1 the factor k_2 was uniquely determined by $k_1 \pmod{S_1}$. This implies that

$$hk_1h^{-1} \equiv k_1 \pmod{S_1}$$

for every h, and hence k_1 belongs to the center of H/S_1. We have therefore:

THEOREM 3. *For $m=2$ the normal subgroups of the symmetry which are contained in the basis group can be obtained by the following construction. Two normal subgroups $K\supset S_1$ of H are chosen such that K/S_1 belongs to the center of H/S_1. Furthermore T is some fixed automorphism of order 2 of K/S_1. Then the normal subgroups $\overline{N}$ consist of all those elements*

$$\kappa = [k_1, k_2]$$

where k_1 and k_2 run through K subject to the condition that

$$k_2 \equiv k_1^T \pmod{S_1}.$$

In the general case where $m \geqq 3$ the elements of $\overline{N}$ are given by (14). Here again the last factor is uniquely determined by the $m-1$ first ones. Let us transform κ by $\eta=[h, 1, \cdots, 1]$. Then

$$\eta\kappa\eta^{-1} = [h^{-1}kr_1h, kr_2, \cdots, kr_{m-1}, k^T(r_1 \cdots r_{m-1})^{-1}].$$

This shows first that

$$h^{-1}kr_1h = ks$$

where s belongs to S_2, and since the last factor must be the same as before one finds $s=r_1$; and hence kr_1 belongs to the center of H.

We have therefore:

THEOREM 4. *For $m \geqq 3$ the normal subgroups of $\Sigma_m(H)$ contained in the basis group are obtained by the construction in Theorem 2 with the additional condition that $K \supset S_2 \supset S_1$ are normal subgroups of H such that K/S_1 belongs to the center of H/S_1.*

5. **Other normal subgroups of $\Sigma_m(H)$.** We shall complete our investigations of the normal subgroups of the symmetry by the determination of those normal subgroups which are not contained in the basis group.

Let $\overline{M}$ be such a normal subgroup of Σ_m. The cross-cut

$$\overline{N} = \overline{M} \cap V_m$$

consists of the multiplications contained in $\overline{M}$ and $\overline{N}$ is normal in Σ_m. Since $\overline{N}$ is contained in V_m it must have one of the forms determined in the preceding. But in this case one can make a further statement about $\overline{N}$.

Let α be a substitution in $\overline{M}$ and

$$\gamma = \begin{pmatrix} x_1, & x_2, & \cdots, & x_n \\ c_1x_2, & c_2x_3, & \cdots, & c_nx_1 \end{pmatrix}$$

a cycle in α, while

$$\eta = [h_1, \cdots, h_m]$$

is an arbitrary multiplication in V_m. Since $\overline{M}$ is normal in Σ_m it must contain the multiplication defined by the commutator

$$\eta\alpha\eta^{-1}\alpha^{-1} = [\cdots, c_n^{-1}h_n^{-1}c_nh_1, c_1^{-1}h_1^{-1}c_1h_2, \cdots]$$

where only the factors corresponding to the cycle γ have been indicated. Since the elements h_i in H are arbitrary it follows that the factor at any position can be made to take an arbitrary value. Using these results we can say:

THEOREM 5. *Let $\overline{M}$ be a normal subgroup of Σ_m not contained in the basis group. The multiplications contained in $\overline{M}$ form a normal subgroup $\overline{N}$ of Σ_m in which $H=K$, i.e., the factors in any fixed position run through the whole group and the quotient H/S_1 for $\overline{N}$ is an Abelian group.*

Again we shall have to consider the case $m=2$ separately. In this case every element which is not a multiplication is a cycle of length 2 and every element in $\overline{M}$ must belong to one of the cosets

(15) $$\bar{M} = \bar{N} + \tau\bar{N}$$

where τ is any fixed cycle,

(16) $$\tau = \begin{pmatrix} x_1, & x_2 \\ x_2, & ax_1 \end{pmatrix}$$

in $\bar{M}$. Since

$$\tau^2 = [a, a]$$

must belong to $\bar{N}$ it follows from Theorem 3 that

(17) $$a^T \equiv a \pmod{S_1},$$

i.e., a is an element in H invariant under the automorphism T of H/S_1.

Now let us consider a group $\bar{M}$ defined by (15) where τ is the cycle (16) and $\bar{N}$ is a normal subgroup of Σ_2 contained in V_2 having the properties indicated in Theorem 5. If the condition (17) is satisfied it is obvious that the elements defined by (15) form a group. In order that it should be normal in Σ_2 it is necessary and sufficient that it be transformed into itself by any multiplication

$$\kappa = [h_1, h_2]$$

and by the transposition $\sigma = (1, 2)$. Since $\bar{N}$ already is a normal subgroup of Σ_m it is sufficient to show that τ is transformed into an element of $\bar{M}$. This in turn is equivalent to the fact that the commutators

$$\kappa\tau\kappa^{-1}\tau^{-1}, \qquad \sigma\tau\sigma^{-1}\tau^{-1}$$

shall belong to $\bar{N}$.

For the first commutator one finds

$$\kappa\tau\kappa^{-1}\tau^{-1} = [a^{-1}h_2^{-1}ah_1, h_1^{-1}h_2].$$

This implies that

$$a^{-1}h_2^{-1}ah_1 \equiv (h_1^{-1}h_2)^T \pmod{S_1}$$

and since H/S_1 is Abelian

$$(h_1^{-1}h_2)^T \equiv (h_1^{-1}h_2)^{-1} \pmod{S_1},$$

hence T is an automorphism of H/S_1 taking each element into its inverse.

For the second commutator one finds

$$\sigma\tau\sigma^{-1}\tau^{-1} = [a^{-1}, a]$$

which is always in $\bar{N}$. By the condition (17) one must also have $a^2 = 1$.

THEOREM 6. *The normal subgroups of $\Sigma_2(H)$ not contained in $V_2(H)$ can*

be obtained by the following construction. A normal subgroup S_1 of H is chosen so that H/S_1 is Abelian. Then the group $\overline{N}$ consisting of the elements

$$\kappa = [k_1, k_2]$$

where

$$k_2 \equiv k_1^{-1} \pmod{S_1}$$

forms a normal subgroup of Σ_2 contained in V_2. The general normal subgroup $\overline{M}$ of Σ_2 not in V_2 consists of the two cosets

$$\overline{M} = \overline{N} + \tau\overline{N}$$

where

$$\tau = \begin{pmatrix} x_1, & x_2 \\ x_2, & ax_1 \end{pmatrix}$$

with

$$a^2 \equiv 1 \pmod{S_1}.$$

We shall now turn to the general case where $m \geqq 3$. As before we denote by $\overline{M}$ a normal subgroup of Σ_m not contained in V_m and by $\overline{N}$ the cross-cut of $\overline{M}$ with V_m. Furthermore we denote by P the subgroup of $\overline{M}$ consisting of permutations only. Since $\overline{M}$ is normal it follows that P is normal in the symmetric group S_m.

We shall prove first:

THEOREM 7. *Every normal subgroup $\overline{M}$ of Σ_m, $m \geqq 3$, not contained in V_m contains permutations.*

Proof. Let α be a substitution in $\overline{M}$ and γ a cycle in α of length $n \geqq 2$. According to a previous result α can always be so transformed that

$$\gamma = \begin{pmatrix} x_1, & \cdots, & x_{n-1}, & x_n \\ x_2, & \cdots, & x_n, & ax_1 \end{pmatrix}.$$

We shall now have to separate several cases. Let us assume first $n \geqq 3$. Then one puts $\tau = (1, 2)$ and one finds that

$$\tau\alpha\tau^{-1}\alpha^{-1} = (1, 2, 3)$$

belongs to P.

We can assume therefore that every cycle which occurs in any substitution in $\overline{M}$ has the lengths 1 and 2. If α contains at least two cycles of length 2 one can write

$$\alpha = \begin{pmatrix} x_1, & x_2 \\ x_2, & ax_1 \end{pmatrix}\begin{pmatrix} x_3, & x_4 \\ x_4, & bx_3 \end{pmatrix} \cdots .$$

If one puts $\tau=(1, 3)$ one obtains

$$\tau\alpha\tau^{-1}\alpha^{-1} = (1, 3)(2, 4).$$

Finally if α contains only one cycle of length 2 one can write

$$\alpha = \begin{pmatrix} x_1, & x_2, & x_3, & \cdots \\ x_2, & ax_1, & bx_3, & \cdots \end{pmatrix}.$$

When $\tau=(1, 3)$ one finds

$$\tau\alpha\tau^{-1}\alpha^{-1} = \begin{pmatrix} x_1, & x_3, & x_2 \\ x_3, & b^{-1}x_2, & bx_1 \end{pmatrix}$$

and $\overline{M}$ contains a cycle of length 3 contrary to assumption.

Since P is a normal subgroup of the symmetric group, there exist only the following three possibilities:

1. P is the symmetric group S_m.
2. P is the alternating group A_m.
3. When $m=4$, P can be the four group P_4 consisting of the elements

$$1, (1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3).$$

On the basis of this result one can prove a stronger statement than Theorem 5:

THEOREM 8. *Let $\overline{M}$ be a normal subgroup of $\Sigma_m(H)$, $m\geqq 3$, not contained in the basis group. Then the multiplications in $\overline{M}$ form a normal subgroup $\overline{N}$ consisting of the elements*

$$\eta = [h_1, \cdots, h_m] \tag{18}$$

where the h_i run through H subject to the condition

$$h_1 \cdots h_m \equiv 1 \pmod{S_1}. \tag{19}$$

Here S_1 is a normal subgroup of H such that H/S_1 is Abelian.

Proof. Let us assume first that the group of permutations P in $\overline{M}$ is the symmetric or alternating group. Then it contains the cycle $\tau=(1, 2, 3)$, and hence also the commutator

$$\kappa\tau\kappa^{-1}\tau^{-1} = [k_3^{-1}k_1, k_1^{-1}k_2, k_2^{-1}k_3, 1, \cdots, 1] \tag{20}$$

where

$$\kappa = [k_1, \cdots, k_m]$$

is an arbitrary multiplication. But in (20) two factors may be made perfectly arbitrary. Hence when one compares it with the group in Theorem 2 one must have $S_2=H$ and an element (18) in $\overline{N}$ is subject only to condition (19).

In the exceptional case when $P=P_4$ is the four group, it must contain $(1, 2)(3, 4)=\tau$ and

$$\kappa\tau\kappa^{-1}\tau^{-1} = [k_2^{-1}k_1,\ k_1^{-1}k_2,\ k_4^{-1}k_3,\ k_3^{-1}k_4,\ 1, \cdots, 1]$$

belongs to $\overline{N}$. Here the third and the first factors may take arbitrary values in H and the same argument as before applies.

We may also observe that in Theorem 8 the quotient group $V_m/\overline{N}$ is Abelian and this holds even when $m=2$.

We shall now proceed to the actual construction of the normal subgroups of Σ_m. The quotient group Σ_m/V_m is isomorphic to the symmetric group. Furthermore

$$\overline{M} \cup V_m/V_m \simeq \overline{M}/\overline{N}$$

is a group isomorphic to a normal subgroup of S_m, and hence the group $\overline{M}/\overline{N}$ is isomorphic to either S_m, A_m or P_4.

The group P was also isomorphic to one of these groups. Let us now suppose first that $P\simeq\overline{M}/\overline{N}$ and hence $\overline{M}=P\cup\overline{N}$. We then have the following theorem:

THEOREM 9. *Let P be a normal subgroup of the symmetric group and $\overline{N}$ a normal subgroup of the basis group of the type defined in Theorem 8. Then*

$$\overline{M} = P \cup \overline{N}$$

is a normal subgroup of Σ_m.

Proof. The group $\overline{N}$ is normal in Σ_m by assumption, and hence it is only necessary to show that an element in P is transformed into an element in $\overline{M}$. Since P is normal in S_m it is sufficient to show that a permutation π in P is transformed into an element in $\overline{M}$ by any multiplication

$$\kappa = [k_1, \cdots, k_m].$$

But if one forms the commutator

$$\kappa\pi\kappa^{-1}\pi^{-1} = [\cdots, k_{\pi^{-1}i}^{-1}k_i, \cdots]$$

one sees, since H/S_1 is Abelian, that the condition (19) is satisfied.

It remains therefore only to consider the case where P is isomorphic to a proper subgroup of $\overline{M}/\overline{N}$. This can only happen in the following two cases:

1. $P=P_4$ and $\overline{M}/\overline{N}\simeq S_m$ or $\overline{M}/\overline{N}\simeq A_m$.
2. $P=A_m$ and $\overline{M}/\overline{N}\simeq S_m$.

The first case is easily disposed of. If $\overline{M}/\overline{N}$ is isomorphic to S_m or A_m it follows that $\overline{M}$ must contain a cycle of length 3 and by the argument used in the proof of Theorem 7 it follows that P contains a cycle $(1, 2, 3)$ contrary to the assumption that $P=P_4$.

We consider therefore the second case where $P=A_m$. Then there must exist some permutation

$$\lambda = \begin{pmatrix} x_1, & x_2, & x_3, & \cdots \\ a_1x_2, & a_2x_1, & a_3x_3, & \cdots \end{pmatrix}$$

in $\overline{M}$. According to Theorem 8 the factors in any element in $\overline{N}$ may be taken arbitrarily in H except for one of them and it is therefore no limitation to assume $a_3 = \cdots = a_m = 1$. Furthermore one sees that λ may be so transformed that

$$\lambda = \begin{pmatrix} x_1, & x_2 \\ x_2, & ax_1 \end{pmatrix}. \tag{21}$$

Since λ^2 must belong to $\overline{N}$ one finds

$$a^2 \equiv 1 \pmod{S_1}. \tag{22}$$

We can then prove:

THEOREM 10. *Let*

$$\overline{M}_A = A_m \cup \overline{N}$$

be a normal subgroup of Σ_m *defined by the procedure of Theorem* 9 *and let* $L = \{\lambda\}$ *be the cyclic subgroup defined by a cycle* (21), *where* a *satisfies the condition* (22). *Then*

$$\overline{M} = L \cup \overline{M}_A$$

is a normal subgroup of Σ_m.

Proof. As before it is sufficient to prove that the commutator of λ with a permutation and with a multiplication again belongs to $\overline{M}$ and this is easily verified.

We can now summarize the preceding investigation as follows:

For $m=2$ the normal subgroups of the symmetry are of the forms given in Theorem 3 and Theorem 6.

For $m \geqq 3$ the normal subgroups of the symmetry are of the forms determined by Theorem 4, Theorem 9 and Theorem 10.

6. **Commutator groups.** We shall now determine the commutator group Σ_m' of the symmetry $\Sigma_m(H)$. The commutator group of H shall be denoted by H'. It is then easily seen that the commutator group V_m' of the basis group consists of all elements

$$\eta' = [h_1', \cdots, h_m']$$

where the h_i' run through the elements of H' independently.

From the well known properties of the symmetric group it follows that

the commutator group of the symmetric group is equal to the alternating group A_m for all m, when one puts $A_2=1$ in the special case $m=2$. Thus the commutator group Σ' must contain A_m.

Next let us form the commutator of the two elements (x_1, x_2) and $[h, 1, \cdots, 1]$ where h is an arbitrary element in H. One finds that this commutator is the multiplication

$$[h^{-1}, h, 1, \cdots, 1].$$

From this result one concludes as in §3 that Σ' also contains the group $V_m^*(H)$ consisting of all multiplications

$$\eta = [h_1, \cdots, h_m] \tag{23}$$

where

$$h_1h_2 \cdots h_m \equiv 1 \pmod{H'}. \tag{24}$$

The quotient group V/V^* is seen to be Abelian and isomorphic to the group H/H^*.

From these remarks one can conclude:

THEOREM 11. *The commutator group of the symmetry is*

$$\Sigma_m'(H) = A_m \cup V_m^*$$

where A_m is the alternating group and V_m^* the normal subgroup of Σ_m contained in V_m consisting of the elements (23) *which satisfy the condition* (24).

Proof. It has already been shown that the commutator group must contain A_m and V_m^*. Thus it is sufficient to show that the quotient group $\Sigma_m/A_m \cup V_m^*$ is Abelian. Obviously any two multiplications commute (mod V_m^*) since this group contains V_m'. Furthermore the quotient group is seen to be generated by residue classes defined by multiplications and the transposition (x_1, x_2) and these generating classes are also seen to commute (mod $A_m \cup V_m^*$).

Next let us determine the second commutator group $\Sigma_m''(H)$. Since the alternating group is simple, one has $A_m' = A_m$ except for the well known exceptions,

$$A_2' = 1, \qquad A_3' = 1, \qquad A_4' = P_4.$$

These remarks show that for $m \geqq 5$ the group Σ_m'' must contain A_m except for $m=4$ when it contains P_4.

We shall now determine certain multiplications which must be contained in Σ''. Let us suppose that $m \geqq 3$.

If h_1 and h_2 are arbitrary elements in H then Σ' contains the multiplications

$$\eta_1 = [h_1, h_1^{-1}, 1, \cdots, 1], \qquad \eta_2 = [h_2, 1, h_2^{-1}, 1, \cdots, 1].$$

By forming the commutator it follows that Σ'' contains

$$\eta_1\eta_2\eta_1^{-1}\eta_2^{-1} = [h_1h_2h_1^{-1}h_2^{-1}, 1, \cdots, 1],$$

hence Σ'' must contain V_m'.

We shall now prove that in most cases Σ'' contains the group V_m^*. To obtain this result let $n \geqq 3$ be some odd number not exceeding m. The cycle

$$\gamma = \begin{pmatrix} x_1, & \cdots, & x_{m-1}, & x_m \\ x_2, & \cdots, & x_m, & x_1 \end{pmatrix}$$

then belongs to A_m. Furthermore let

$$\eta = [h_1, \cdots, h_n, 1, \cdots, 1]$$

with

$$h_1h_2 \cdots h_n \equiv 1 \pmod{H'} \tag{25}$$

be an element in V_m^*. The commutator

$$\pi = \eta^{-1}P^{-1}\eta P = [h_2h_1^{-1}, h_3h_2^{-1}, \cdots, h_nh_{n-1}^{-1}, h_1h_n^{-1}, 1, \cdots, 1]$$

must then belong to Σ''. We shall consider $\pi \pmod{V_m'}$ and eliminate h_n by means of the condition (25). Since H/H' is Abelian one finds

$$\pi \equiv [h_2h_1^{-1}, h_3h_2^{-1}, \cdots, h_1^{-1} \cdots h_{n-2}^{-1}h_{n-1}^{-2}, h_1^2h_2 \cdots h_{n-1}, 1, \cdots, 1] \pmod{V_m'} \tag{26}$$

and here the h_i $(i=1, \cdots, n-1)$ can take on arbitrary values in H.

For $h_1=1$ the first multiplier in (26) becomes the arbitrary element h_2. Thus the group of multiplications in Σ'' must have the property that the multiplier at any position runs through H.

Next let us make all h_i $(i=1, 2, \cdots, n-1)$ equal to the same element h. From (26) one obtains

$$\pi \equiv [1, \cdots, 1, h^{-n}, h^n, 1, \cdots, 1] \pmod{V_m'}$$

and one sees that Σ'' contains

$$\eta_n = [h^{-n}, h^n, 1, \cdots, 1]$$

for any odd n such that $m \geqq n \geqq 3$.

If now $m \geqq 5$ one can take $n=3$ and $n=5$ and by combining the results one finds that Σ'' contains

$$\eta_1 = [h^{-1}, h, 1, \cdots, 1],$$

hence it contains the full group V_m^*.

In the special case $m=4$ one can obtain the same result as follows. The alternating group A_4 contains the substitution

$$\gamma_1 = \begin{pmatrix} x_1, & x_2 \\ x_2, & x_1 \end{pmatrix}\begin{pmatrix} x_3, & x_4 \\ x_4, & x_3 \end{pmatrix}$$

and one finds by the same process of forming commutators that Σ_4'' must contain η_4 and since it contains η_3 it also contains η_1. Thus it has been shown:

THEOREM 12. *For $m \geqq 5$ the second commutator group $\Sigma_m''(H)$ is equal to the first commutator group $\Sigma_m'(H)$. For $m=4$ one has*

$$\Sigma_4''(H) = P_4 \cup V_4^*$$

where V_4^* consists of all multiplications

$$\eta = [h_1, h_2, h_3, h_4]$$

with

$$h_1 h_2 h_3 h_4 \equiv 1 \pmod{H'}.$$

In the case $m=3$ the second commutator group is contained in the basis group, hence it must be a normal subgroup of the form determined by Theorem 9. We mention the result without proof:

THEOREM 13. *For $m=3$ the second commutator group $\Sigma_3''(H)$ is a subgroup of the basis group consisting of the multiplications $\eta = [h_1, h_2, h_3]$ where the factors are subject to the conditions*

$$h_1 \equiv hk_1^3, \; h_2 \equiv hk_2^3, \; h_3 \equiv h^{-2}(k_1k_2)^{-3} \pmod{H'}$$

and h, k_1, k_2 are arbitrary elements in H.

One finally finds:

THEOREM 14. *For $m=2$ the second commutator group $\Sigma_2''(H)$ consists of the multiplications*

$$\eta = [h_1, h_2]$$

where h_1 and h_2 run through H' subject to the condition

$$h_2 \equiv h_1 \pmod{\overline{H}}$$

where $\overline{H}$ is the group generated by the commutators of an element in H with an element in H'.

For $m=2, 3, 4$ one can also determine the higher commutator groups. One can also, through similar methods, determine the commutator groups of the normal subgroups of the symmetry.

CHAPTER III. AUTOMORPHISMS

1. The basis group as characteristic subgroup. We shall now turn to the problem of the determination of the automorphisms of the symmetry. Again the case $m=2$ presents a real exception and we shall assume, therefore, to begin with that $m\geqq 3$. We shall assume also that the group H with respect to which the symmetry $\Sigma_m(H)$ is constructed is a *finite* group. This condition is, however, not essential in most of the results, but it simplifies several of the special considerations which are necessary for the lowest values of m.

We begin our investigations by deducing the following theorem:

THEOREM 1. *For $m\geqq 3$ the basis group is a characteristic subgroup of the symmetry.*

Proof. Let us assume that the theorem is not true. Then there would exist some automorphism α which takes a subgroup $\overline{M}$ into the basis group V_m. The group $\overline{M}$ is a normal subgroup of Σ_m not contained in the basis group and hence its form is determined by Theorem 6, Chapter II.

The quotient group Σ_m/V_m is isomorphic to the symmetric group S_m. Since $\overline{M}$ is transformed into V_m by α one must also have

$$\Sigma_m/\overline{M}\simeq S_m\simeq\Sigma_m/V_m. \tag{1}$$

Next let us write

$$\overline{K}=\overline{M}\cup V_m,\qquad \overline{N}=\overline{M}\cap V_m.$$

Since $\overline{K}$ is a normal subgroup of Σ_m containing V_m there exist only the following three possibilities

$$\overline{K}=S_m\cup V_m,\qquad \overline{K}=A_m\cup V_m,\qquad \overline{K}=P_4\cup V_4, \tag{2}$$

where as before A_m denotes the alternating group and P_4 the four group.

According to (1) the two quotient groups Σ_m/V_m and $\Sigma_m/\overline{M}$ are isomorphic. Thus the quotient groups $\overline{K}/V_m$ and $\overline{K}/\overline{M}$ must have the same order and since both of them are isomorphic to a normal subgroup of S_m one must have

$$\overline{K}/\overline{M}\simeq\overline{K}/V_m.$$

Then one concludes from the second law of isomorphism

$$\overline{K}/V_m\simeq\overline{M}/\overline{N}\simeq V_m/\overline{N}. \tag{3}$$

The form of the normal subgroup $\overline{N}$ can also be determined from Theorem 6, Chapter II. According to this theorem there must exist a normal subgroup S of H such that H/S is Abelian. Then $\overline{N}$ consists of all multiplications

$$\eta=[h_1,\cdots,h_m]$$

where the elements h_i run through H subject only to the condition

$$h_1 \cdots h_m \equiv 1 \pmod{S}.$$

This shows that the quotient group $V_m/\overline{N}$ is isomorphic to H/S, hence S is a proper subgroup of H. But furthermore, since $V_m/\overline{N}$ is Abelian it follows from (2) and (3) that one is led to a contradiction in all cases except possibly in the two special cases

$$m = 3, \quad \overline{M}/\overline{N} \simeq A_3,$$
$$m = 4, \quad \overline{M}/\overline{N} \simeq P_4.$$

To prove the theorem also for these two cases we proceed somewhat differently. Since the two groups $\overline{M}$ and V_m are isomorphic, their commutator groups $\overline{M}'$ and V_m' and also the corresponding quotient groups $\overline{M}/\overline{M}'$ and V_m/V_m' must be isomorphic.

The group V_m' is seen to consist of the elements

$$(4) \qquad \kappa = [k_1, \cdots, k_m]$$

where the k_i run through the commutator group H' of H. We can now show that in the case $m \geqq 3$ the commutator group of the cross-cut $\overline{N}$ of $\overline{M}$ and V_m consists of the same elements (4). When h_1 and h_2 are arbitrary elements in H the group $\overline{N}$ must contain the two elements

$$[h_1, h_1^{-1}, 1, \cdots, 1], \qquad [h_2, 1, h_2^{-1}, 1, \cdots, 1];$$

thus the commutator group $\overline{N}'$ must contain

$$[h_1h_2h_1^{-1}h_2^{-1}, 1, \cdots, 1]$$

and consequently $\overline{N}' = V_m'$.

We obtain the desired contradiction by showing that $\overline{M}'$ contains V_m' as a proper subgroup. According to Theorem 7, Chapter II, there exists some permutation in $\overline{M}$ and the proof of this theorem shows that there must exist some permutation P in $\overline{M}$ taking x_1 into x_2. But then for any element

$$\eta = [h_1, \cdots, h_m]$$

the commutator

$$\eta^{-1}P^{-1}\eta P = [h_2h_1^{-1}, \cdots]$$

must belong to $\overline{M}'$. By a suitable choice of h_1 and h_2 one obtains that in $\overline{M}'$ the factors of any variable in a multiplication can run through the full group H. But according to the construction of $\overline{N}$ the group

$$S \supset H'$$

is a proper subgroup of H hence $\overline{M}'$ contains V_m' as a proper subgroup.

The result obtained in Theorem 1 can be supplemented as follows:

THEOREM 2. *In the case $m=2$ the basis group need not be a characteristic subgroup of the symmetry.*

To prove this statement we turn to the simplest example of a symmetry where H is a cyclic group of order 2 generated by an element a for which $a^2=1$. Then the group $\Sigma_2(H)$ consists of the following eight elements:

$$\begin{gathered} 1,\ [a, a],\ [a, 1],\ [1, a], \\ (x_1, x_2),\ (x_1, x_2)[a, a],\ (x_1, x_2)[a, 1],\ (x_1, x_2)[1, a]. \end{gathered} \tag{5}$$

If one puts $x=(x_1, x_2)$, $y=(x_1, x_2)[a, 1]$, then the group $\Sigma_2(2)$ can also be defined by the generating relations

$$x^2 = 1, \qquad y^4 = 1, \qquad xyx^{-1} = y^3. \tag{6}$$

We shall now determine the group of outer automorphisms of this group. The two first elements in (5) constitute the center of the group and they are, therefore, invariant by all automorphisms. Since the two last elements in (5) are the only ones of order 4, any automorphism α must take y into y or y^3. But in the latter case one can, according to (6), multiply α by the inner automorphism x and one can, therefore, always assume

$$y^\alpha = y. \tag{7}$$

The element x must be transformed under α into some other element of order 2. It cannot be left invariant because then according to (7) every element would be invariant. It cannot be transformed into $[a, a]$ since this element belongs to the center. There remain, therefore, only the three possibilities

$$x^\alpha = [a, 1], \qquad x^\alpha = [1, a], \qquad x^\alpha = (x_1, x_2)[a, a]. \tag{8}$$

The last alternative in (8) can be ruled out since one finds that it corresponds to the inner automorphism defined by y. One finds that the two remaining correspondences (8) differ only by an inner automorphism and that they actually define an outer automorphism

$$\begin{aligned} (x_1, x_2)^\alpha &= [a, 1], & [a, 1]^\alpha &= (x_1, x_2)[a, a], \\ ((x_1, x_2)[a, a])^\alpha &= [1, a], & [1, a]^\alpha &= (x_1, x_2). \end{aligned} \tag{9}$$

Since this automorphism does not leave the basis group invariant, Theorem 2 has been proved.

2. **The automorphisms of the basis group.** Before we proceed to the determination of the automorphisms of the symmetry it is necessary to make some remarks about the automorphisms of the basis group. This is of course the same as the determination of the automorphisms of a direct product of m factors all isomorphic to H.

Let H be some group. An endomorphism of H is a homomorphic image of H upon some subgroup. Since any homomorphic image is isomorphic to

some quotient group H/A all endomorphisms are obtainable from the subgroups of H isomorphic to some such quotient group.

The endomorphisms of the basis group $V_m(H)$ can easily be determined from those of the group H. The group $V_m(H)$ is the direct product of the m isomorphic groups $\overline{H}^{(i)}$ each consisting of the elements

$$\eta_i = [1, \cdots, 1, h_i, 1, \cdots, 1]. \tag{10}$$

Now let T be some endomorphism of $V_m(H)$. Then one must have

$$\eta_i^T = [h_i^{T_1^{(i)}}, \cdots, h_i^{T_m^{(i)}}] \tag{11}$$

where the $T_s^{(i)}$ denote correspondences in H. One sees immediately that these correspondences must be endomorphisms in H. Furthermore, since the two elements η_i and η_j are permutable one finds that

$$h^{T_s^{(i)}} k^{T_s^{(j)}} = k^{T_s^{(j)}} h^{T_s^{(i)}} \qquad (s = 1, 2, \cdots, m),\ i \neq j, \tag{12}$$

for arbitrary elements h and k in H.

Through the correspondences (11) one finds the element corresponding to the general element

$$\eta = [h_1, \cdots, h_m] \tag{13}$$

to be

$$\eta^T = [h_1^{T_1^{(1)}} h_2^{T_1^{(2)}} \cdots h_m^{T_1^{(m)}},\ h_1^{T_2^{(1)}} h_2^{T_2^{(2)}} \cdots h_m^{T_2^{(m)}},\ \cdots]. \tag{14}$$

Conversely one sees that if a set of endomorphisms

$$T_s^{(i)} \qquad (s = 1, 2, \cdots, m;\ i = 1, 2, \cdots, m) \tag{15}$$

are given and the element η^T corresponding to η in (13) is given by (14) then T represents an endomorphism of $V_m(H)$ if the conditions (12) are satisfied.

THEOREM 3. *All endomorphisms of $V_m(H)$ are obtainable through the possible sets of m^2 endomorphisms* (15) *of H satisfying* (12), *by letting the general element η in* (13) *correspond to the element* (14).

The endomorphism thus defined is an automorphism if and only if the set of equations

$$h_1^{T_i^{(1)}} h_2^{T_i^{(2)}} \cdots h_m^{T_i^{(m)}} = k_i \qquad (i = 1, 2, \cdots, m)$$

have a unique solution $h_1, h_2, \cdots, h_m$ for arbitrary $k_1, k_2, \cdots, k_m$ in H.

3. **Conditions on automorphisms.** We shall now proceed to the actual determination of the outer automorphisms of the symmetry in the general case.

These investigations are based upon the following remarks. Let G be a group and N some normal subgroup of G such that G splits over N

$$G = M \cup N; \qquad M \cap N = E. \tag{16}$$

Now let M' and N' be groups isomorphic to M and N under the isomorphisms α and β, respectively. Then to each element $\mu \cdot \nu$ in G can be made to correspond a unique symbol $\mu^\alpha \cdot \nu^\beta$. Now let us determine when this correspondence defines an isomorphism between G and a group

$$G' = M' \cup N', \qquad M' \cap N' = E.$$

Since the multiplication in G is defined by the rule

$$\mu_1\nu_1 \cdot \mu_2\nu_2 = \mu_1\mu_2 \cdot \overset{-1}{\mu_2}\, \nu_1\mu_2 \cdot \nu_2$$

one must have

$$\overset{\alpha}{\mu_1}\overset{\beta}{\nu_1} \cdot \overset{\alpha}{\mu_2}\overset{\beta}{\nu_2} = (\mu_1\mu_2)^\alpha \cdot (\overset{-1}{\mu_2}\, \nu_1\mu_2)^\beta \cdot \overset{\beta}{\nu_2}.$$

This reduces to the necessary and sufficient condition

$$\mu^\alpha \nu^\beta (\mu^{-1})^\alpha = (\mu\nu\mu^{-1})^\beta \tag{17}$$

which must hold for all μ in M and ν in N.

For the symmetry one can always write according to §1, Chapter I,

$$\Sigma_m = S_m \cup V_m, \qquad S_m \cap V_m = E \tag{18}$$

where S_m is the symmetric group and V_m the basis group. This relation corresponds to the preceding (16). Now let τ be any automorphism of Σ_m. Since V_m is a characteristic subgroup for $m \geqq 3$ according to Theorem 1, one obtains from (18)

$$\Sigma_m = S_m^\tau \cup V_m, \qquad S_m^\tau \cap V_m = E.$$

The possible form of such a group S_m^τ is determined by Theorem 10, Chapter I. There exists an isomorphism τ between S_m and S_m^τ generated by the correspondence

$$(1, i)^\tau = [1, a_{1,i}, \cdots, a_{m,i}](1, i) \tag{19}$$

which indicates the elements corresponding to the transpositions $(1, i)$ generating S_m. Here the $a_{i,j}$ are elements of H generating a subgroup of H homomorphic to the symmetric group on $m-1$ letters by means of the correspondence

$$(i, j) \to a_{i,j}, \qquad i \neq 1, j \neq 1.$$

We have now the same situation as in the introductory remarks. There exists an isomorphism α between S_m and S_m^τ, hence τ and α can only differ in their effect upon S_m by an automorphism of S. But according to a result due

to *Hölder*[9], the symmetric group on m letters has only inner automorphisms except when $m=6$. Since we are only interested in outer automorphisms of Σ_m we can, therefore, when $m\neq 2$ and $m\neq 6$, multiply τ by such an inner automorphism that the isomorphism α between S_m and S_m^{τ} is given by the relations (19).

Next let the $T_s^{(i)}$ in (15) denote a set of endomorphisms of H satisfying the permutability conditions (12). The effect of any automorphism β of V_m upon any element η in V_m given by (13) is then expressed in (14), where it must be additionally verified that the β thus defined is an automorphism and not only an endomorphism.

After having defined the correspondences α through (19) and β through (14), one can find all automorphisms of Σ_m through the combination of them in the manner indicated in the preceding. It remains only to verify when the conditions corresponding to (17) are satisfied for the two correspondences. One finds easily that it is sufficient to consider these conditions for generating elements of the two groups, hence for $\mu=(1,\, i)$, $\nu=\eta_j$ where η_j is defined by (10).

From

$$(1,\, i)\eta_j(1,\, i)^{-1} = \eta_j, \qquad i \neq j,\ j \neq 1, \tag{20}$$

follows first

$$[1, a_{2,i}, \cdots, a_{m,i}](1,\, i)[h_j^{T_1^{(j)}}, \cdots, h_j^{T_m^{(j)}}](1,\, i)[1, a_{2,i}^{-1}, \cdots, a_{m,i}^{-1}] = [h_j^{T_1^{(j)}}, \cdots, h_j^{T_m^{(j)}}].$$

After simplification this leads to the relations

$$\begin{aligned} h^{T_i^{(j)}} &= h^{T_1^{(j)}}, \\ a_{k,i} h^{T_k^{(j)}} a_{k,i}^{-1} &= h^{T_k^{(j)}} \qquad (k = 1, 2, \cdots, m) \end{aligned} \tag{21}$$

which hold for an arbitrary h in H. The first relation (21) shows that if one puts

$$T^{(j)} = T_j^{(j)}, \qquad R^{(j)} = T_i^{(j)}, \qquad i \neq j, \tag{22}$$

then one has the simpler form for the automorphism β when $j\neq 1$:

$$\eta_j^{\beta} = [h_j^{R^{(j)}}, \cdots, h_j^{R^{(j)}}, h_j^{T^{(j)}}, h_j^{R^{(j)}}, \cdots, h_j^{R^{(j)}}]. \tag{23}$$

With the notations (22) the second relation (21) splits into the two

$$\begin{aligned} a_{k,i} h^{R^{(j)}} a_{k,i}^{-1} &= h^{R^{(j)}}, \qquad k \neq j, \\ a_{j,i} h^{T^{(j)}} a_{j,i}^{-1} &= h^{T^{(j)}}. \end{aligned} \tag{24}$$

---

(9) O. Hölder, *Bildung zusammengesetzter Gruppen*, Mathematische Annalen, vol. 46 (1895), pp. 321–422.

Next we turn to the relation corresponding to (20) for $j=1$ or $j=i$,

$$(1, i)\eta_1(1, i) = \eta_j.$$

This leads by (17) to the following three relations deduced in the same manner as (21)

$$h^{T_i^{(1)}} = h^{R^{(i)}}, \qquad h^{T_1^{(1)}} = h^{T^{(i)}}, \qquad h^{T_k^{(1)}} = a_{k,i}h^{R^{(i)}}a_{k,i}^{-1}. \tag{25}$$

From the second of these relations it follows that

$$T = T_1^{(1)} = T^{(i)} \tag{26}$$

is independent of i. From the first relation (25) one finds supplementing (23),

$$\eta_1^{\beta} = [h_1^{T}, h_1^{R^{(2)}}, \cdots, h_1^{R^{(m)}}] \tag{27}$$

and by combination of (23) and (27) one finds for the general element η in V_m defined in (13)

$$\begin{aligned}\eta^{\beta} = [h_1^{T}h_2^{R^{(2)}} \cdots h_m^{R^{(m)}}, h_1^{R^{(2)}}h_2^{T}h_3^{R^{(3)}} \cdots h_m^{R^{(m)}},\\ \cdots, h_1^{R^{(m)}}h_2^{R^{(2)}} \cdots h_{m-1}^{R^{(m-1)}}h_m^{T}].\end{aligned} \tag{28}$$

By combining the first and the last relations (25) one also has

$$h^{R^{(i)}} = a_{i,j}h^{R^{(j)}}a_{i,j}^{-1} \tag{29}$$

and according to (24) $a_{i,j}$ is permutable with any h^T and any $h^{R^{(k)}}$ if $i \neq k, j \neq k$.

After having derived the conditions corresponding to (17) it remains only to verify that the correspondence β defined by (23) and (27) or (28) actually is an automorphism of the basis group. We observe first that the permutability conditions (12) in this case became

$$\begin{aligned} h^T k^{R^{(i)}} &= k^{R^{(i)}} h^T,\\ h^{R^{(i)}} k^{R^{(j)}} &= k^{R^{(j)}} h^{R^{(i)}}, \qquad i \neq j.\end{aligned} \tag{30}$$

When these relations hold the correspondence β is an endomorphism according to Theorem 3. To make it an automorphism the relations

$$\begin{aligned} h_1^{T}h_2^{R^{(2)}}h_3^{R^{(3)}} \cdots h_m^{R^{(m)}} &= k_1,\\ h_1^{R^{(2)}}h_2^{T}h_3^{R^{(3)}} \cdots h_m^{R^{(m)}} &= k_2,\\ \cdots\cdots\cdots\cdots\cdots&,\\ h_1^{R^{(m)}}h_2^{R^{(2)}} \cdots h_{m-1}^{R^{(m-1)}}h_m^{T} &= k_m\end{aligned} \tag{31}$$

must have a unique set of solutions h_i for any set of elements k_i in H. To

analyze these conditions we put

$$h_i = y_i h_1, \qquad k_i = x_i k_1 \tag{32}$$

and form the quotient between the first and the ith relation (31). Using the permutability conditions (30) this leads to the $m-1$ relations

$$y_i^T (y_i^{-1})^{R^{(i)}} = x_i. \tag{33}$$

If one puts symbolically

$$h^{T-R^{(i)}} = h^T (h^{-1})^{R^{(i)}}, \tag{34}$$

then $T-R^{(i)}$ is a correspondence of H to a subset of itself. It should be noted that this correspondence need not be an endomorphism. When the notation (34) is used one can write (33) slightly simpler,

$$y_i^{T-R^{(i)}} = x_i \qquad (i = 2, \cdots, m). \tag{35}$$

Since these conditions must be satisfied by some y_i for every set of elements x_i in H, it follows that $T-R^{(i)}$ is a one-to-one correspondence of H to itself and conversely if this is the case the y_i are uniquely determined by the x_i.

When (32) is substituted in the relations (31) it follows by the permutability conditions (30) and from (35) that the relations (31) reduce to a single condition

$$y_2^{R^{(2)}} \cdots y_m^{R^{(m)}} h_1^T h_1^{R^{(2)}} \cdots h_1^{R^{(m)}} = k_1. \tag{36}$$

Here it is convenient to put

$$h^S = h^{T+R^{(2)}+\cdots+R^{(m)}} = h^T h^{R^{(2)}} \cdots h^{R^{(m)}},$$

and it is seen immediately that

$$S = T + R^{(2)} + \cdots + R^{(m)}$$

is an endomorphism of H. If, however, there shall exist a unique solution h_1 of (36) for given k_1 and $y_2, \cdots, y_m$ it is seen that S must be an automorphism of H. Conversely if this is the case and if all $T-R^{(i)}$ are one-to-one correspondences of H to itself, the correspondence β is an automorphism of the basis group. This concludes the set of conditions for an automorphism of Σ_m.

4. **Explicit determination of the automorphisms.** Through an analysis of the previous necessary and sufficient conditions for an automorphism one can obtain an explicit representation of all automorphisms of the symmetry.

Let us observe first that since the correspondence $T-R^{(i)}$ is a one-to-one correspondence of H to itself, every element $\bar{h}$ in H can be written in the form

$$\bar{h} = h^T (h^{-1})^{R^{(i)}} \tag{37}$$

where h is some element in H. Let H^T and $H^{R^{(i)}}$ denote the subgroups on

which H is mapped by these endomorphisms. Then one concludes from (37) that

$$H = H^T \cup H^{R^{(i)}}.$$

But if $i \neq j$ then the elements of $R^{(j)}$ are permutable with the elements of $H^{R^{(i)}}$ and H^T hence with all elements of H. This shows that $H^{R^{(j)}}$ belongs to the center of H. But since $R^{(i)}$ and $R^{(j)}$ differ only by an inner automorphism they must be identical and

$$R = R^{(i)} \tag{38}$$

does not depend on i, and one can write

$$H = H^T \cup H^R.$$

The elements a_{ij} are permutable with the elements of H^T and also with the elements of H^R since they belong to the center.

This shows that also all a_{ij} belong to the center. But the a_{ij} are known to generate a group A homomorphic to the symmetric group on $m-1$ letters. Thus one has only the two possibilities

$$A \simeq 1, \qquad A \simeq S_{m-1}/A_{m-1}.$$

In the first case $a_{ij}=1$ for all i and j. In the second case A is cyclic of order 2. Thus one can put in both cases

$$a_{ij} = a, \qquad i \neq j, \qquad a^2 = 1 \tag{39}$$

where a belongs to the center of H.

The endomorphism R projects H upon a subgroup of its center. One sees that such an endomorphism can only exist if the anti-center, i.e., the quotient group of the commutator group of H has a subgroup isomorphic to a subgroup of the center. One sees also that if R has this property and if the a_{ij} are defined through (39) then all the preceding conditions are satisfied except that it remains to verify that the correspondence β in (28) actually is an automorphism.

Under the conditions just derived the relation (28) simplifies to

$$\eta^\beta = [h_1^S p^R, h_2^S p^R, \cdots, h_m^S p^R] \tag{40}$$

where we have put

$$S = T - R \tag{41}$$

and

$$p = h_1 h_2 \cdots h_m. \tag{42}$$

It should be noted that since H^R belongs to the center the correspondence S in (41) is actually an automorphism of H.

The correspondence (40) can obviously be considered the product of two correspondences of the form

$$\eta^{\sigma} = [h_1^S, h_2^S, \cdots, h_m^S],$$
$$\eta^{\kappa} = [h_1 p^K, h_2 p^K, \cdots, h_m p^K]$$

where S and p are defined by (41) and (42) while $K = RS^{-1}$ is also an endomorphism projecting H upon a subgroup of the center. Since σ is an automorphism of V_m the correspondence β is an automorphism if and only if κ has this property. By a previous method one sees that this is the case if and only if the correspondence

$$h \to h(h^m)^K = h^{1+mK}$$

is an automorphism of H. This last automorphism is a central automorphism where every element is multiplied by an element in the center.

We have thus finally arrived at the main result:

THEOREM 4. *Let $m \neq 2$ and $m \neq 6$. Any automorphism of the symmetry $\Sigma_m(H)$ can be obtained by transformation from an automorphism τ constructed in the following manner:*

The images of the permutations in Σ_m are determined by

$$(1, i)^{\tau} = [1, a, \cdots, a, 1, a, \cdots, a](1, i),$$
$$(i, j)^{\tau} = [a, \cdots, a](i, j), \qquad i \neq 1, j \neq 1,$$

where a is any element of order 2 belonging to the center of H. The image of an element in the basis group

$$\eta = [h_1, h_2, \cdots, h_m]$$

is

$$\eta^{\tau} = [(h_1 p^K)^S, (h_2 p^K)^S, \cdots, (h_m p^K)^S]$$

where S is any automorphism of H and

$$p = h_1 h_2 \cdots h_m,$$

while K is any endomorphism of H projecting H upon a subgroup of its center in such a manner that $1 + mK$ is a central automorphism of H.

From this result all outer automorphisms of $\Sigma_m(H)$ can be determined. Let us find the conditions for $\Sigma_m(H)$ to be a complete group, i.e., a group in which all automorphisms are inner automorphisms. First it is easily observed that the correspondence determined in Theorem 4 for the permutations in Σ_m cannot be obtained by an inner automorphism except when $a = 1$. Thus the center of H can contain no elements of order 2. When $a = 1$ all permutations in Σ_m are left invariant by τ and the only inner automorphisms in Σ_m which

will have this property are the scalars $[h]$. Next let us choose K such that $h^K=1$ for every h in H. Then one finds

$$\eta^\tau = [h_1^S, \cdots, h_m^S]$$

and this can only be obtained by transformation with a scalar if S is an inner automorphism of H. Finally a correspondence

$$\eta^\tau = [h_1 p^K, \cdots, h_m p^K]$$

is never obtainable by transformation with a scalar except when $h^K=1$ for every h. Thus we have shown:

THEOREM 5. *Let $m\neq 2$ and $m\neq 6$. The necessary and sufficient condition for the symmetry $\Sigma_m(H)$ to be complete is that H be complete and have a center of odd order. Furthermore, there shall exist no endomorphism K of H such that H^K belongs to the center and $1+mK$ is an automorphism of H.*

If H has no center $\Sigma_m(H)$ is complete if and only if H is complete.

Theorems 4 and 5 give the main results on the automorphisms of the symmetry when $m\neq 2$ and $m\neq 6$. The case $m=2$ will be considered in the following. The case $m=6$ is complicated by the fact that in this case the group S_m has an outer automorphism. But since this automorphism is known(4), the preceding method can be used to determine the automorphisms of Σ_m also in this case. Since the calculations are somewhat laborious they have not been carried through here. It would be of interest, if someone would take the trouble of carrying through this investigation.

5. **Automorphisms for $m=2$.** We shall now analyze the automorphisms of the symmetry in the exceptional case $m=2$. As we have already observed in §1, the basis group $V_2(H)$ need not be invariant by the automorphisms of $\Sigma_2(H)$. We shall say, however, that an automorphism is regular if it does have the property that it leaves $V_2(H)$ invariant. We shall show a little later that except for a special type of groups H the automorphisms of $\Sigma_2(H)$ are always regular.

We shall first determine the form of the regular automorphisms of $\Sigma_2(H)$ by the same method as the one used in §3 in the general case. According to Theorem 2, Chapter I, the symmetry splits regularly over its basis group and it can be assumed, therefore, corresponding to (19) that we have multiplied the regular automorphism τ in question by an inner automorphism such that

$$(1, 2)^\tau = (1, 2).$$

---

(4) According to O. Hölder (loc. cit.) the single outer automorphism to be considered may be taken as

$$(1, 2) \to (1, 2)(3, 4)(5, 6), \quad (1, 3) \to (1, 6)(2, 3)(4, 5), \quad (1, 4) \to (1, 5)(2, 4)(3, 6),$$
$$(1, 5) \to (1, 3)(2, 5)(4, 6), \quad (1, 6) \to (1, 4)(2, 6)(3, 5).$$

Furthermore one must have

$$[h_1, 1]^\tau = [h_1^{T_1^{(1)}}, h_1^{T_2^{(1)}}][1, h_2]^\tau = [h_2^{T_1^{(2)}}, h_2^{T_2^{(2)}}]$$

with the permutability conditions (12). By taking the image of the relation

$$(1, 2)[h, 1](1, 2) = [1, h]$$

one obtains the conditions

$$T = T_1^{(1)} = T_2^{(2)}, \qquad R = T_2^{(1)} = T_1^{(2)}$$

and consequently

$$[h_1, h_2]^\tau = [h_1^T h_2^R, h_2^T h_1^R]$$

where the elements of the two groups H^T and H^R are permutable. It remains only to determine when this homomorphism of $V_2(H)$ is an automorphism. One finds that the relations

$$h_1^T h_2^R = k_1, \qquad h_2^T h_1^R = k_2$$

can be solved uniquely for h_1 and h_2 with arbitrary k_1 and k_2 in H if and only if $T-R$ is a one-to-one correspondence of H to itself while $T+R$ is an automorphism of H. Thus we have shown:

THEOREM 6. *For $m=2$ a regular automorphism of $\Sigma_2(H)$ differs only by an inner automorphism from the automorphism defined by*

$$(1, 2)^\tau = (1, 2),$$
$$[h_1, h_2]^\tau = [h_1^T h_2^R, h_2^T h_1^R]$$

where the endomorphisms T and R of H have the property that

$$h_1^T h_2^R = h_2^R h_1^T$$

and $T+R$ is an automorphism of H while $T-R$ is a one-to-one correspondence of H to itself.

The existence of two such endomorphisms T and R of H implies a permutable decomposition

$$H = H^T \cup H^R.$$

Conversely if a permutable decomposition of H exists then one can obtain by a few additional conditions that endomorphisms T and R with the desired properties will exist. We shall not discuss these conditions here.

We shall now determine when irregular automorphisms of $\Sigma_2(H)$ can exist. For an irregular automorphism τ there must exist some subgroup $\overline{M}$ such that

$$\overline{M}^{\tau} = V_2(H).$$

Here $\overline{M}$ must be normal and obviously

$$\Sigma_2(H) = \overline{M} \cup V_2(H).$$

Since $\overline{M}$ is isomorphic to V_2 one can write

$$\overline{M} = \overline{H} \times \overline{\overline{H}} \tag{43}$$

where each direct component is isomorphic to H. The group $\overline{M}$ must contain cycles

$$\gamma = \begin{pmatrix} x_1, & x_2 \\ ax_2, & bx_1 \end{pmatrix} \tag{44}$$

and consequently at least one of the groups $\overline{H}$ and $\overline{\overline{H}}$ contains such cycles. Since $\overline{M}$ is normal in Σ_2 and since any transformation of $\overline{M}$ results in some other decomposition (43) we can assume that $\overline{H}$ contains a cycle

$$\gamma_0 = \begin{pmatrix} x_1, & x_2 \\ x_2, & ax_1 \end{pmatrix} \tag{45}$$

in the normal form.

The centralizer of such a cycle (45) has already been determined in §4, Chapter I and it was found to consist of the elements of the form

$$\gamma_0^i[c, c] \qquad (i = 0, 1)$$

where c is any element in the centralizer of a in H.

The order of the centralizer of γ_0 in Σ_2 is, therefore, $2c_a$ where c_a is the order of the centralizer of a in H. On the other hand, all elements in $\overline{\overline{H}}$ must belong to the centralizer of γ_0 according to the direct decomposition (43). Furthermore some elements of $\overline{H}$, for instance the powers of γ_0 must belong to it. Thus the order of the centralizer becomes at least $n_0 \cdot n_H$ where n_0 is the order of γ_0 and n_H the order of H. By comparison with the previous determination of the order of the centralizer this can only be possible if $c_a = n_H$ and $n_0 = 2$. But then

$$\gamma_0^2 = [a, a] = 1$$

and $a = 1$, so that γ_0 in (45) takes the special form $\gamma_0 = (x_1, x_2)$. It has been shown, therefore, that every cycle γ in (44) contained in $\overline{H}$ has the order 2 and consequently it has the special form

$$\gamma = \begin{pmatrix} x_1, & x_2 \\ ax_2, & a^{-1}x_1 \end{pmatrix}.$$

Now let us assume again that $\overline{H}$ has been so transformed that it contains

$\gamma_0=(x_1, x_2)$. We expand $\overline{H}$ in cosets

$$\overline{H} = \overline{H}_0 + \gamma_0\overline{H}_0$$

where $\overline{H}_0$ is the normal subgroup of $\overline{H}$ consisting of its multiplications. If $[a, b]$ is an element of $\overline{H}_0$ then $\gamma_0[a, b]$ is a cycle contained in $\overline{H}$ and from the preceding result it follows that $b=a^{-1}$. Thus all the elements of $\overline{H}_0$ have the form

$$\alpha = [a, a^{-1}].$$

If

$$\beta = [b, b^{-1}]$$

is another element in $\overline{H}_0$ then their product

$$\alpha\beta = [ba, b^{-1}a^{-1}] = [ba, (ba)^{-1}]$$

is also in $\overline{H}_0$. This implies $ab=ba$ and $\overline{H}_0$ is Abelian. Since $\overline{H}$ is isomorphic to H this shows that H is obtained from an Abelian group H_0 by an extension of order 2.

One can make a further statement about the group H_0. It follows from the preceding that the group $\overline{H}$ cannot contain any other elements of the centralizer of γ_0 than 1 and γ_0 itself. However all elements $[a, a]$ belong to the centralizer of γ_0 and such elements must be contained in $\overline{H}$ if $\overline{H}_0$ contains an element $[a, a^{-1}]$ with $a^2=1$. This occurs if and only if $\overline{H}_0$ contains elements of order 2. Hence we have shown:

THEOREM 7. *Irregular automorphisms of $\Sigma_2(H)$ can only exist if H contains a normal Abelian subgroup H_0 of odd order and index* 2.

If H has the form indicated in Theorem 7 one can write

$$H = A \cup H_0, \qquad A \cap H_0 = E$$

where

$$A = \{a\}, \qquad a^2 = 1$$

is a cyclic group of order 2. Let us write $V_2(H_0)$ for the group consisting of all elements

$$[h_1^{(0)}, h_2^{(0)}]$$

where $h_1^{(0)}$ and $h_2^{(0)}$ run through H_0 independently. Obviously $V_2(H_0)$ is a normal subgroup of $\Sigma_2(H)$. Since its order is an odd number it is relatively prime to its index 8, consequently $V_2(H_0)$ is a characteristic subgroup. It is also seen that $\Sigma_2(H)$ splits over $V_2(H_0)$:

$$\Sigma_2(H) = \Sigma_2(A) \cup V_2(H_0), \qquad \Sigma_2(A) \cap V_2(H_0) = E \tag{46}$$

where $\Sigma_2(A)$ is the group of order (8) whose elements are given in (5). We have shown that $\Sigma_2(A)$ was the simplest group for which irregular automorphisms existed.

The automorphisms of $\Sigma_2(H)$ can now be constructed by the method indicated in §3 from the decomposition (46). The possible automorphisms of $V_2(H_0)$ are defined by

$$[k, 1]^\tau = [k_1^{T_1^{(1)}}, k_2^{T_2^{(1)}}][1, k_2]^\tau = [k_2^{T_1^{(2)}}, k_2^{T_2^{(2)}}]$$

where k_1 and k_2 are elements of H_0 and the T denote endomorphisms of H_0 satisfying the permutability conditions (12).

Since the order of $V_2(H_0)$ is relatively prime to its order the group $\Sigma_2(H)$ splits regularly over $V_2(H_0)$, i.e., any other group which can replace $\Sigma_2(A)$ in (46) is obtained from $\Sigma_2(A)$ by an inner automorphism. Since τ may be multiplied by an arbitrary inner automorphism one can assume that τ leaves $\Sigma_2(A)$ invariant. This means that the 8 elements of $\Sigma_2(A)$ are transformed by τ according to the automorphism of $\Sigma_2(A)$ already determined in §1. The elements

$$x = (x_1, x_2), \qquad y = (x_1, x_2)[a, 1]$$

generate the group and one has according to (7) and (9)

$$x^\tau = [a, 1], \qquad y^\tau = y.$$

The method indicated in §3 can now be used. We shall not go into the details of the calculations, but only give the final result:

THEOREM 8. *Let H be a group for which $\Sigma_2(H)$ can have irregular automorphisms. Then H has a normal Abelian subgroup H_0 of odd order and of index 2, hence*

$$H = \{a\} \cup H_0, \qquad a^2 = 1.$$

The irregular automorphisms of $\Sigma_2(H)$ are then determined by

$$(x_1, x_2)^\tau = [a, 1],$$
$$[a, 1]^\tau = (x_1, x_2)[a, a], \qquad [1, a]^\tau = (x_1, x_2),$$

and

$$[h_1, h_2]^\tau = [(h_1 h_2^a)^T, (h_1 h_2)^T], \qquad h_1, h_2 \text{ in } H_0,$$

where $h^a = aha^{-1}$ and T denotes an automorphism of H_0 such that

$$(h^a)^T = (h^T)^a.$$

Furthermore the element a must satisfy the condition that every element in H_0 is expressible as a commutator $hah^{-1}a^{-1}$.

Chapter IV. Monomial representations

1. Construction of monomial representations. Let G be some group. A *monomial representation* of G is a homomorphism of G to some subgroup M_G of the symmetry $\Sigma_m(H)$ of some group H. A monomial representation may also be considered a homomorphism of G to a set of matrices such that to each g in G there corresponds a monomial matrix M_g, i.e., a matrix of the special type that each line and column contains only one nonvanishing element. Since the multiplication of such matrices involves only the products of the elements which occur in them, one can suppose that the elements in the monomial matrices belong to some group H.

If K is some normal subgroup of H then any monomial representation in H also gives a monomial representation in the quotient group H/K simply by considering the elements of H (mod K).

For the monomial groups one can introduce the analogues of the ordinary concepts of transitivity and primitivity. Any monomial representation can be written uniquely as a product of transitive representations.

All transitive monomial representations of a group can be obtained by the following construction:

Let H be a subgroup of G of finite index m and let

$$G = Hg_1 + \cdots + Hg_m, \qquad g_1 = 1, \tag{1}$$

be the corresponding coset expansion. Then for any z in G one has a relation

$$g_i z = h_i^{(z)} g_{z:i}. \tag{2}$$

This relation defines for each z a monomial substitution

$$M_z = \begin{pmatrix} x_1, & x_2, & \cdots, & x_m \\ h_1^{(z)} x_{z:1}, & h_2^{(z)} x_{z:2}, & \cdots, & h_m^{(z)} x_{z:m} \end{pmatrix}. \tag{3}$$

The elements $h_i^{(z)}$ in H shall be called the *factors* of z. If t is another element in G and

$$g_j t = h_j^{(t)} g_{t:j},$$

then

$$g_i(zt) = h_i^{(z)} g_{z:i} t = h_i^{(z)} h_{z:i}^{(t)} g_{tz:i},$$

and one sees that

$$M_{zt} = M_t M_z.$$

This shows that there exists a transposed homomorphism between the elements in G and the monomial substitutions (3). One can of course obtain a direct homomorphism by letting M_z in (3) correspond to the element z^{-1}.

If K is any normal subgroup of H then one obtains a monomial representation of G with respect to H/K by considering the factors $h_i^{(z)}$ in (3) (mod K).

From the manner in which the construction of the monomial representation has been performed it follows that there exist substitutions M_z taking an arbitrary variable x_i into any hx_j. The elements in H correspond to substitutions for which $x_1 \rightarrow hx_1$ and those in K to substitutions in which $x_1 \rightarrow x_1$.

Conversely it is easily seen that all transitive monomial representations may be obtained in this manner. For a given monomial group, the group H consists of those substitutions for which $x_1 \rightarrow h_1 x_1$ and K those for which $x_1 \rightarrow x_1$, while the generator g_i in (1) is taken as the substitution for which $x_1 \rightarrow x_i$.

The preceding theory applies only to a group G having a subgroup of finite index. There is however no serious difficulty in extending the theory to the case of an arbitrary transfinite number of cosets in (1). This would involve the extension of the theory of the symmetry $\Sigma_m(H)$ to such transfinite degrees.

The question immediately arises when the monomial representation is isomorphic to the group G. If an element z shall correspond to the unit substitution, one must have, according to (2), $g_i z g_i^{-1} \subset K$ for all i, hence K contains a normal subgroup of G. Conversely all the elements of such a subgroup of K must correspond to the unit substitution.

We may summarize these results as follows:

THEOREM 1. *All transitive monomial representations of a group G are obtainable by selecting a subgroup H of G and a normal subgroup K of H and constructing the monomial representation of G in H/K by the coset expansion* (1) *and defining the factors of the substitutions from* (2). *The corresponding representation is isomorphic to G/N where N is the largest normal subgroup of G contained in K*[6].

2. **Various properties of the monomial representations.** We observe first that the monomial representation of a group obtained by the process just indicated depends on the choice of the representatives g_i in the coset expansion (1). If g_i is replaced by

$$\bar{g}_i = \bar{h}_i g_i$$

then one finds from (2)

$$\bar{g}_i z = \bar{h}_i g_i z = \bar{h}_i h_i^{(z)} g_{z:i} = \bar{h}_i h_i^{(z)} \bar{h}_{z:i}^{-1} \bar{g}_{z:i},$$

hence the factors in the new representation are

$$f_i = \bar{h}_i h_i^{(z)} \bar{h}_{z:i}^{-1}.$$

But this shows that M_z is replaced by the substitution

(6) See W. K. Turkin, loc. cit.

$$\overline{M}_z = \kappa^{-1} M_z \kappa$$

where κ is the multiplication

$$\kappa = [\bar{h}_1, \cdots, \bar{h}_m].$$

One could also have changed the order of the cosets in (1) and this would have the effect of renumbering the variables x_i. But any such change can also be obtained by transforming M_z by a permutation and we can therefore say:

THEOREM 2. *All the different transitive monomial representations of a group G in quotient group H/K which can be obtained through permutations of cosets and change of representatives are conjugate subgroups of $\Sigma_m(H/K)$. Conversely any two transitive conjugate subgroups of $\Sigma_m(H/K)$ can be considered as representatives of the same group where one representation is obtained from the other by the two processes indicated.*

The actual determination of the monomial substitution corresponding to a given element z may often be made by such a transformation. One chooses the representatives in the coset expansion (1) in sets

$$g_i, g_i z, \cdots, g_i z^{f_i - 1} \tag{4}$$

where f_i is the smallest exponent such that

$$g_i z^{f_i} \subset H g_i$$

or

$$z^{f_i} \subset g_i^{-1} H g_i.$$

If one writes

$$g_i z^{f_i} = d_i g_i \tag{5}$$

then one finds that in the monomial representation $z \rightarrow M_z$ where M_z is in the normal form

$$M_z = \prod_{i=1}^{r} \begin{pmatrix} x_1, & \cdots, & x_{f_i - 1}, & x_{f_i} \\ x_2, & \cdots, & x_{f_i}, & d_i x_1 \end{pmatrix}. \tag{6}$$

The cycles in this substitution correspond to the terms in the expansion of the group into double cosets

$$G = HZ + Hg_2 Z + \cdots + Hg_m Z \tag{7}$$

where $Z = \{z\}$ is the cyclic group defined by z. Each double coset in (7) contains f_i cosets with respect to H and

$$\sum_{i=1}^{r} f_i = m. \tag{8}$$

These remarks permit us to decide when the substitution M_z corresponding to z can be transformed into a permutation. In this case one must have $d_i=1$ in (6) for every i and one concludes:

THEOREM 3. *Let z be an element in G and M_z the monomial substitution corresponding to z in the representation of G with respect to a quotient group H/K. Then a necessary and sufficient condition that M_z be transformable into a permutation is that the elements of the cyclic group $Z=\{z\}$ and its conjugates which belong to H also belong to K.*

One can also find when the monomial substitution corresponding to z is a multiplication. In this case one must have, for every g_i in (1),

$$g_i z = h_i^{(z)} g_i \tag{9}$$

or

$$g_i z g_i^{-1} \subset H;$$

hence the element z must belong to a normal subgroup of G contained in H. Conversely if z belongs to such a group the relation (9) will hold.

THEOREM 4. *Those elements of G which correspond to multiplications in the monomial representation of G with respect to H/K form a group which is the greatest normal subgroup of G contained in H.*

Let us consider briefly the case when H is normal in G. Then every element in G has the form $z=hg_j$. The monomial substitution corresponding to h is determined by

$$g_i h = h' g_i, \qquad h' = h^{g_i},$$

hence M_h is the multiplication

$$M_h = [h^{g_1}, \cdots, h^{g_m}].$$

Corresponding to the representative g_j in the coset expansion one has

$$g_i g_j = f_{i,j} g_{i,j}$$

where the $f_{i,j}$ form the factor set of the extension G over H, hence

$$M_{g_j} = \begin{pmatrix} x_1, & \cdots, & x_m \\ f_{1,j}x_{1,j}, & \cdots, & f_{m,j}x_{m,j} \end{pmatrix}.$$

3. **The normalizer of a subgroup in the symmetry.** Let G be a group represented monomially with respect to a quotient group H/K. As a subgroup of the complete monomial group $\Sigma_m(H/K)$ the group G will have a certain normalizer N_G. We shall now study the properties of those automorphisms which the elements of N_G induce in the group G.

Let ν be an element of N_G taking the variable x_1 into x_i. One sees then that

$$\nu H \nu^{-1} = H_i \tag{10}$$

where H_i is the subgroup of G consisting of those substitutions multiplying x_i by a factor. This shows that ν transforms all groups H_i among themselves.

Now there also exists some inner automorphism g of G such that

$$gHg^{-1} = H_i$$

and one finds

$$\mu H \mu^{-1} = H, \qquad \mu = \nu g^{-1}.$$

One needs therefore only to study substitutions μ transforming H into itself. But if

$$\mu = \begin{pmatrix} x_1, & \cdots \\ h_1 x_1, & \cdots \end{pmatrix}$$

one can multiply μ further by such an element in M_H that $h_1 = 1$. In this case one sees that

$$\mu H \mu^{-1} = H, \qquad \mu K \mu^{-1} = K$$

and further by transformation with μ the quotient group H/K is left elementwise invariant.

Conversely let A be an automorphism of G leaving H/K elementwise invariant. Then one can associate with A a monomial substitution

$$\mu = \begin{pmatrix} \cdots, & x_i, & \cdots \\ \cdots, & t_i x_{A:i}, & \cdots \end{pmatrix}$$

where the factors t_i are defined by

$$\overset{A}{g_i} = t_i g_{A:i}.$$

Next we form the substitution

$$\mu M_z \mu^{-1} = \begin{pmatrix} \cdots, & x_i, & \cdots \\ \cdots, & t_{A^{-1}:i} h^{(z)}_{A^{-1}:i} t_{zA^{-1}:i} x_{AzA^{-1}:i}, & \cdots \end{pmatrix}$$

for an arbitrary element z in G. We shall show that

$$\mu M_z \mu^{-1} = M_{z^A}.$$

To construct the factors of M_{z^A} we observe that

$$\overset{A}{g_i z} = \overset{-1}{t}_{A^{-1}:i} \overset{A}{(g_{A^{-1}:i} z)} = \overset{-1}{t}_{A^{-1}:i} \overset{A}{(h^{(z)}_{A^{-1}:i} g_{zA^{-1}:i})}$$

and since A leaves H/K invariant

$$g_{i\varepsilon}^{A} = t_{A^{-1};i}h_{A^{-1};i}^{(\varepsilon)}t_{\varepsilon A^{-1};i}g_{A\varepsilon A^{-1};i}$$

as desired. We have proved therefore:

THEOREM 5. *Let G be a group which is represented as a transitive monomial group of degree m with respect to a quotient group H/K. Then the normalizer N_G of G in the symmetry $\Sigma_m(H/K)$ consists of all those substitutions which aside from a factor in G induce such automorphisms in G that the quotient group H/K is left elementwise fixed.*

Under certain circumstances all automorphisms of G may be induced by the elements of the normalizer N_G. Let us determine when this occurs. We denote by A an arbitrary automorphism of G. Then one must have

$$H^A = H_i = H^{g_i}$$

for some g_i in G, hence A has the effect of an inner automorphism upon H and H is a *typical*(⁶) subgroup of G. Then $B = Ag_i^{-1}$ leaves H invariant and

$$K^B = K, \qquad K^A = K^{g_i},$$

hence H and K are compatible typical subgroups. Finally H/K must be left elementwise invariant by B.

THEOREM 6. *The necessary and sufficient condition that all automorphisms of G shall be induced by the elements of the normalizer N_G of G in the symmetry $\Sigma_m(H/K)$ is that H and K be typical compatible subgroups of G such that any automorphism of G which leaves H fixed, shall, aside from an inner automorphism of H, leave the quotient group H/K elementwise fixed.*

The conditions of this theorem are satisfied when:

1. H is typical,
2. K is characteristic in H,
3. H/K is complete, i.e. all automorphisms are inner automorphisms.

The two first conditions are, for instance, always satisfied for a Sylow group and its normalizer.

4. **The centralizer of a subgroup in the symmetry.** Again let G be a group represented as a monomial group with respect to a quotient group H/K. As a subgroup of $\Sigma_m(H/K)$ the group G has a certain centralizer Z_G whose properties we shall now study.

The center $\bar{C}$ of Σ_m consists of the multiplications

$$\mu = [c, \cdots, c]$$

where c belongs to the center of H/K. Obviously the center of Σ_m belongs to Z_G. We can prove further:

(⁶) Such groups will be studied in greater detail in another paper on group theory which will be submitted for publication shortly.

Any multiplication in Σ_m belonging to Z_G must belong to the center $\overline{C}$ of Σ_m.

Proof. Let the multiplication

$$\mu = [c_1, \cdots, c_n]$$

belong to Z_G. Since there exist elements in G carrying x_i into hx_i for any i, where h is an arbitrary element in H/K, one must have

$$c_i h c_i^{-1} = h$$

and c_i must belong to the center of H/K. Furthermore since G contains elements carrying x_1 into x_i one finds $c_1 = c_i$ for every i.

Next we show:

Let σ be a substitution belonging to Z_G which only changes some variable by a factor. Then σ is a multiplication and belongs to the center of Σ_m.

Proof. Let σ take x_i into hx_i and let M_s be some substitution in G carrying x_i into some other variable x_j. Then one finds

$$\sigma : x_j = \sigma M_s : x_i = M_s \sigma : x_i = M_s : hx_i = hx_j$$

and σ is a multiplication. This result also shows that two substitutions in Z_G taking a variable x_i into the same variable x_j differ only by a factor which is an element of $\overline{C}$.

We shall now give a method of constructing elements of Σ_m which belong to Z_G. Let z be an element in G which belongs to the normalizer of H and in addition has the property that it leaves the quotient group H/K elementwise invariant by transformation. Then one can associate with z a unique monomial substitution

$$S_z = \begin{pmatrix} \cdots, & x_i, & \cdots \\ \cdots, & t_i x_{z:i}, & \cdots \end{pmatrix}$$

where

$$zg_i = t_i g_{z:i}.$$

Next let y be an arbitrary element in G and let us consider the product zg_iy. By the associative law this product, which belongs to G can be evaluated in two ways, namely, first

$$(zg_i)y = (t_i g_{z:i})y = t_i h_{z:i}^{(y)} g_{yz:i},$$

and secondly by the stated property of z

$$z(g_i y) = z h_i^{(y)} g_{y:i} = h_i^{(y)} \cdot t_{y:i} g_{zy:i}.$$

This implies, however, $M_y S_z = S_z M_y$, hence S_z belongs to Z_G.

We have thus established a correspondence $z \rightarrow S_z$ between the elements z of G which leave H/K elementwise invariant by transformation and certain elements S_z of Z_G. This correspondence is seen to be a homomorphism and the elements corresponding to the unit element are the elements of K.

Now conversely let

$$S = \begin{pmatrix} x_1, & \cdots \\ h_1 x_i, & \cdots \end{pmatrix}$$

be an element in Z_G. Then there exists an element z in G such that

$$M_z = \begin{pmatrix} x_1, & \cdots \\ h_1 x_i, & \cdots \end{pmatrix}$$

and the element z is determined uniquely (mod K). Let M_h be the substitution corresponding to an arbitrary element h in H. Then

$$M_z M_h M_z^{-1} : x_1 = M_z M_h M_z^{-1} \cdot S^{-1} \cdot M_z : x_1 = M_z S^{-1} M_h : x_1 = M_z S^{-1} : h_1 x_1 = h_1 x_1.$$

This shows that

$$M_z M_h M_z^{-1} \equiv M_h \text{ (mod } M_K),$$

hence M_z and z have the property of leaving H/K invariant by transformation. By this correspondence between S and z the unit element corresponds to the elements in K. We have shown therefore:

THEOREM 7. *The centralizer Z_G of a group G in the symmetry $\Sigma_m(H/K)$ is isomorphic to the quotient group M/K where M consists of the elements in G which have the property of leaving the quotient group H/K elementwise invariant by transformation.*

One may ask finally when the centralizer consists only of elements in G, hence when Z_G is equal to the center of G. Let z be an element in G leaving H/K invariant by transformation. The corresponding substitution S_z belongs to G if there exists an element y in G such that $zg_i \equiv g_i y$ (mod K). Clearly y must belong to the center C_G of G and for $i=1$ one finds $z = y \cdot k$ where k belongs to K. This gives immediately:

THEOREM 8. *The necessary and sufficient condition that the centralizer of G in $\Sigma_m(H/K)$ be the center C_G of G is that*

$$M = C_G \cup K$$

where M denotes the subgroup of G whose elements have the property of leaving the quotient group H/K elementwise invariant by transformation.

YALE UNIVERSITY,
NEW HAVEN, CONN.

THE FINITE DISPLACEMENT OF THIN RODS(1)

BY

G. E. HAY

CONTENTS

1. Introduction. With reference to problems in elasticity, the terms "small displacement," "finite displacement," "small strain" and "finite strain" have the following meanings which will be made more precise later. If the displacement of every particle in a strained elastic body is much smaller than the greatest linear dimension of the body, the displacement is small; otherwise it is finite. If the elongation per unit length of every element in the body is much smaller than unity the strain is small; otherwise it is finite.

Only those problems involving small strain lie within the scope of ordinary mathematical elasticity. In most problems the strain is small only if the displacement is also small. However, in the case of thin rods and plates the displacement may be finite and yet the resulting strain may be small. Such finite displacement of thin rods will be considered in this paper. At a later date it is proposed to deal with the corresponding problem for thin plates.

The usual theory of the finite displacement of thin rods is due chiefly to Kirchhoff: it has been the subject of a number of papers going back nearly a hundred years(2). An account of the essentials of the Kirchhoff theory has

Presented to the Society, April 11, 1941; received by the editors December 4, 1940.

(1) This paper was written originally with the aid of a scholarship from the National Research Council of Canada, and was revised at Armour Institute of Technology, Chicago.

(2) B. de Saint-Venant, Comptes Rendus de l'Académie des Sciences, Paris, vol. 17 (1843), pp. 942–945, 1020–1031; vol. 19 (1844), pp. 36–44, 181–187. J. Binet, ibid., vol. 18 (1844), pp. 1115–1119. Wantzel, ibid., vol. 18 (1844), pp. 1197–1201. G. Kirchhoff, Journal für die reine und angewandte Mathematik (Crelle), vol. 56 (1859), pp. 285–313, or *Vorlesungen über*

been given by Love([3]). This theory appears to be in need of certain improvements which it is the purpose of the present paper to present. First, in order to deduce the equations governing the finite displacement of a thin rod naturally curved, it was necessary to assume that the rod was displaced to its final position through an intermediate stage in which the rod was straight. No such artificiality is required in the present paper. Secondly, approximation was introduced in a somewhat haphazard fashion. In the present paper approximation is introduced systematically in a manner which permits a theoretical solution of the problem to any desired degree of accuracy. The method is that of development in a power series in a small dimensionless parameter ϵ which tends to zero with the sectional area of the rod. This method was used by G. D. Birkhoff and others for thin plates([4]); it has also been used for thin membranes([5]) and for beams undergoing small displacement([6]).

Synopsis. In this paper we consider rods of uniform cross-section with the external forces acting only on the ends.

The method of tensor calculus is used. Solutions are sought for the equations of equilibrium and compatibility by means of a systematic method of approximation. These solutions contain certain arbitrary functions which can be determined by applying the same method of approximation to certain macroscopic equations of equilibrium. These macroscopic equations of equilibrium express the fact that each element of the rod lying between adjacent cross-sections is in equilibrium under the action of forces acting on its ends. Even though the strain is small, in this work we require stress-strain relations which are more accurate than the usual relations. These relations involve five elastic constants whereas the usual relations involve only two.

The method of approximation hinges on a dimensionless parameter ϵ introduced in the following manner. We consider a singly infinite sequence of thin rods with identical lines of centroids of equal lengths, and with cross-

mathematische Physik, Mechanik, 3rd edition, Leipzig, 1883. A. Clebsch, *Theorie der Elasticität*, Leipzig, 1862, Part 4, or *Théorie de l'Elasticité*, Paris, 1881, chap. 4. A. B. Basset, Proceedings of the London Mathematical Society, vol. 23 (1891–1892), pp. 105–127; American Journal of Mathematics, vol. 17 (1895), pp. 281–317.

([3]) A. E. H. Love, *Elasticity*, 4th edition, Cambridge, 1934, chap. 18.

([4]) G. D. Birkhoff, Philosophical Magazine, vol. 43 (1922), pp. 953–962. C. A. Garabedian, these Transactions, vol. 25 (1923), pp. 343–398; Comptes Rendus de l'Académie des Sciences, Paris, vol. 177 (1923), pp. 942–944; vol. 178 (1924), pp. 619–621; vol. 179 (1924), pp. 381–384; vol. 180 (1925), pp. 257–259; vol. 181 (1925), pp. 319–321; vol. 186 (1928), pp. 1518–1520; vol. 195 (1932), pp. 1369–1371. H. W. Sibert, these Transactions, vol. 33 (1931), pp. 329–369. R. Higdon and D. L. Holl, Duke Mathematical Journal, vol. 3 (1937), pp. 18–34.

([5]) J. L. Synge, Philosophical Transactions of the Royal Society, (A), vol. 231 (1933), pp. 435–477; Transactions of the Royal Society of Canada, (3), section III, vol. 31 (1937), pp. 57–81. G. E. Hay, Canadian Journal of Research, (A), vol. 17 (1939), pp. 106–121; vol. 17 (1939), pp. 123–140.

([6]) J. N. Goodier, Transactions of the Royal Society of Canada, (3), section III, vol. 32 (1938), pp. 1–25.

sections having average diameters continuously decreasing to zero. We define ϵ by the relation

$$(1.1) \qquad \epsilon = \frac{\textit{average diameter of cross-section of unstrained rod}}{\textit{length of unstrained rod}};$$

ϵ then varies in the interval $0<\epsilon<\delta$, where δ is much smaller than unity.

Frequently we shall find it convenient to refer to the order of magnitude of certain quantities. By definition, a quantity q is said to have the order of magnitude of ϵ^r, or to be of order ϵ^r, if q/ϵ^r tends to a finite nonzero limit as ϵ tends to zero. Quantities of order ϵ^0 are said to be finite.

Let us now consider the problem as a whole. A rod of given geometrical form is subjected at one end to a given force-system (force and couple). It takes up some equilibrium configuration under the action of the given force-system and a suitable force-system applied to the other end. Thus the problem may be considered as characterized by

(a) the geometrical form of the unstrained rod,

(b) the force-system applied to one end.

When these are given, the problem is to determine the geometrical form of the unstrained rod by solving the equations of equilibrium and compatibility together with the macroscopic equations of equilibrium. It is found that—

(a) When the applied force-system is of order ϵ^2 or lower, the strain is finite.

(b) When the applied force-system is of order ϵ^5 or higher the displacement is small.

(c) When the applied force-system is of order ϵ^3 or ϵ^4, the displacement is in general finite, while the strain is either finite or small depending on the configuration of the unstrained rod and on the individual components of the applied force-system. We are interested only in those problems in which the displacement is finite and the strain small. It is found that all problems in which this is the case fall naturally into three classes which will now be discussed in turn.

In problems of Class (i) the unstrained rod has in general an arbitrary finite twist and curvature. The given force-system applied to one end is of order ϵ^4 but is otherwise arbitrary. In the strained rod the twist and curvature are in general finite and can be found by the first approximation (Kirchhoff theory) while the elongation per unit length is of order ϵ^2 and can be found only by recourse to the second approximation.

In problems of Class (ii) the unstrained rod has again in general an arbitrary finite twist and curvature, but the given force-system applied to one end of the rod consists of an arbitrary force of order ϵ^3 (much larger than in Class (i)), an arbitrary twisting couple of order ϵ^4, and a particular bending couple of order ϵ^4. The strained rod is nearly straight, with its line of centroids coinciding approximately with the line of action of the given force applied

to the end. In the strained rod the twist (which is finite), the curvature (of order ϵ) and the elongation per unit length (of order ϵ) can all be found by the first approximation. When the magnitude of the curvature is reduced in this way from order ϵ^0 (finite) to order ϵ (small), the rod will be referred to as "straightened." An example of such straightening is worked out in §13.

Class (iii) is a particular case of Class (i). It arises when the form of the unstrained rod is such that two special conditions ((10.20) or (10.21)) are satisfied and the given terminal force-system consists of an arbitrary force and a particular bending couple both of order ϵ^4, the bending couple being so chosen that the strained rod is "straightened." In the strained rod the twist is the same as in the unstrained state, while the curvature (of order ϵ) and the elongation per unit length (of order ϵ^2) can be found only by recourse to the second approximation. A problem of Class (iii) is worked out in §13.

2. **The unstrained rod.** A mathematical investigation involving rods is greatly facilitated by the use of such concepts as "normal cross-section" and "line of centroids." Since these are rather difficult to define for a general curved rod, the present investigation will be limited to those rods which, in the unstrained state, can be defined in the following manner. In Figure 1a,

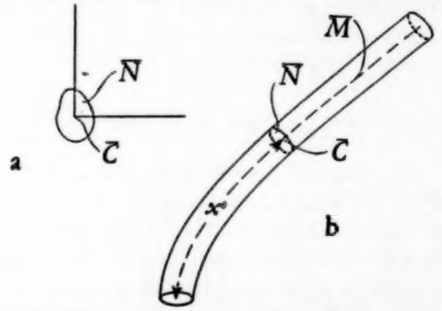


FIG. 1. The unstrained rod: $\overline{N}$ is a cross-section of the rod, $\overline{C}$ is the centroid of $\overline{N}$, and $\overline{M}$ is the line of centroids

$\overline{N}$ is a region bounded by a general closed plane curve, and $\overline{C}$ is the centroid of $\overline{N}$. In Figure 1b, $\overline{M}$ is a general regular skew curve. The rod is generated by $\overline{N}$ when $\overline{C}$ moves along $\overline{M}$ with the plane of $\overline{N}$ normal to $\overline{M}$, as shown in Figure 1b. When a rod is defined in this manner, it has a uniform normal cross-section $\overline{N}$ and a line of centroids $\overline{M}$. Since the rod is to be thin, it is certainly necessary that

$$(\textit{maximum diameter of } \overline{N})/(\textit{length of } \overline{M}) < 1.$$

Curvilinear coordinates x^i ($i=0, 1, 2$) will be used: x^0 is the arc length measured along $\overline{M}$ from one end of the rod to $\overline{C}$: x^1 and x^2 are rectangular cartesian coordinates in the cross-section through $\overline{C}$, with origin at $\overline{C}$ and

the principal axes of inertia of the cross-section as axes. Latin indices will have the range 0, 1, 2; Greek indices the range 1, 2; and, in accordance with the usual convention, repeated indices will indicate summation over the appropriate ranges. The line element in the rod will then be of the form

$$d\bar{s}^2 = \bar{g}_{ij}dx^i dx^j, \tag{2.1}$$

where $\bar{g}_{ij}$ is the metric tensor and $\bar{g}_{ij} = \bar{g}_{ji}$. Because of the nature of the coordinate system,

$$(\bar{g}_{00})_{\bar{C}} = 1, \qquad (\bar{g}_{0\alpha})_{\bar{C}} = 0, \qquad \bar{g}_{\alpha\beta} = \delta^\alpha_\beta, \tag{2.2}$$

where δ^α_β is the Kronecker delta and $(\bar{g}_{ij})_{\bar{C}}$ means that $\bar{g}_{ij}$ is to be evaluated at the point $\bar{C}$ on $\bar{M}$ (where $x^\alpha = 0$).

At each point $\bar{C}$ on $\bar{M}$ there is an orthogonal triad associated with the coordinate axes. To describe the configuration of $\bar{M}$, we shall use a *rotation vector* $\bar{\omega}^i$ specifying the rotation of this triad as $\bar{C}$ moves along $\bar{M}$: $\bar{\omega}^i\bar{\lambda}_i$ is then the rotation per unit length about the direction of a unit vector $\bar{\lambda}_i$, and $\bar{\omega}^0$ and $\bar{\omega}^\alpha$ will be referred to respectively as the twist and curvature of the unstrained rod. The use of $\bar{\omega}^\alpha$ rather than the ordinary components of curvature brings a greater symmetry into subsequent equations.

The remainder of this section is devoted to the derivation of (2.8), which expresses $\bar{\omega}^i$ in terms of $\bar{g}_{ij}$.

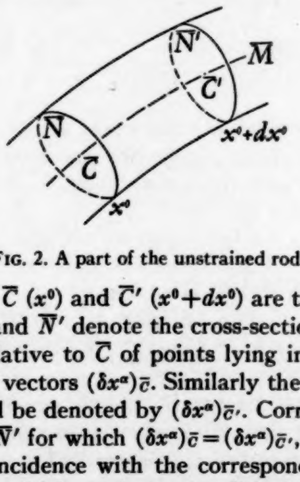


FIG. 2. A part of the unstrained rod

Two adjacent points $\bar{C}$ (x^0) and $\bar{C}'$ (x^0+dx^0) are taken on the line of centroids $\bar{M}$ (Figure 2): $\bar{N}$ and $\bar{N}'$ denote the cross-sections at $\bar{C}$ and $\bar{C}'$, respectively. The positions relative to $\bar{C}$ of points lying in $\bar{N}$ and near $\bar{C}$ will be denoted by infinitesimal vectors $(\delta x^\alpha)_{\bar{C}}$. Similarly the positions relative to $\bar{C}'$ of points lying in $\bar{N}'$ will be denoted by $(\delta x^\alpha)_{\bar{C}'}$. Corresponding to any point in $\bar{N}$ there is a point in $\bar{N}'$ for which $(\delta x^\alpha)_{\bar{C}} = (\delta x^\alpha)_{\bar{C}'}$, and all the points in $\bar{N}$ can be brought into coincidence with the corresponding points in $\bar{N}'$ by a rigid-body displacement, which can be accomplished in two steps: (i) a rigid-body translation in which $\bar{C}$ moves into coincidence with $\bar{C}'$ and $\bar{N}$ moves into a new position $\bar{N}''$; (ii) a rigid-body rotation in which each point in $\bar{N}''$ moves into coincidence with the corresponding point in N'.

In step (i), each vector $(\delta x^\alpha)_{\bar{C}}$ is propagated parallelly from $\bar{C}$ to $\bar{C}'$. The x^0-component of all vectors in the field produced by this parallel propagation vanishes in general only at $\bar{C}$. Presently, differentiation will be performed with respect to the components at $\bar{C}$ of this vector field. Thus it is necessary to introduce a third component $(\delta x^0)^-$ which will be allowed to vanish after the differentiation has been performed. As a result of step (i), $(\delta x^i)_{\bar{C}}$ becomes an infinitesimal vector in $\bar{N}''$ at $\bar{C}'$ with components

$$(2.3)\qquad (\delta x^i)_{\bar{C}} - (\bar{F}^i_{j0})_{\bar{C}}(\delta x^j)_{\bar{C}}dx^0,$$

where $\bar{F}^i_{jk}$ is the Christoffel symbol of the second kind in $\bar{g}_{ij}$,

$$(2.4)\qquad \bar{F}^i_{jk} = \frac{1}{2}\bar{b}^{il}\left(\frac{\partial \bar{g}_{kl}}{\partial x^j} + \frac{\partial \bar{g}_{lj}}{\partial x^k} - \frac{\partial \bar{g}_{jk}}{\partial x^l}\right):$$

$\bar{b}^{ij}$ is the minor of $\bar{g}_{ij}$ in the determinant $|\bar{g}_{ij}|$ divided by that determinant, and $(\bar{F}^i_{jk})_{\bar{C}}$ represents $\bar{F}^i_{jk}$ evaluated at $\bar{C}$ (where $x^\alpha=0$).

As a result of step (ii), (2.3) becomes $(\delta x^i)_{\bar{C}'}$, and since $(\delta x^i)_{\bar{C}}=(\delta x^i)_{\bar{C}'}$, the infinitesimal displacement produced by the rigid-body rotation in step (ii) is

$$(2.5)\qquad \bar{u}^i = (\bar{F}^i_{j0})_{\bar{C}}(\delta x^j)_{\bar{C}}dx^0.$$

This rigid-body rotation is specified by the skew-symmetric tensor

$$(2.6)\qquad \bar{\Omega}_{ij} = \frac{1}{2}\left(\frac{\partial \bar{u}_j}{\partial(\delta x^i)_{\bar{C}}} - \frac{\partial \bar{u}_i}{\partial(\delta x^j)_{\bar{C}}}\right) = \frac{1}{2}dx^0\left(\frac{\partial \bar{g}_{0j}}{\partial x^i} - \frac{\partial \bar{g}_{0i}}{\partial x^j}\right).$$

It can also be specified by the contravariant vector $\bar{\omega}^i dx^0$, such that $\bar{\lambda}_i\bar{\omega}^i dx^0$ is the angle of rotation about a direction specified by a unit vector $\bar{\lambda}_i$. If $\bar{g}=|\bar{g}_{ij}|$ and c^{ijk} are the permutation symbols(7), then

$$(2.7)\qquad \bar{\omega}^i dx^0 = \frac{1}{2}(\bar{g}^{-1/2})_{\bar{C}}c^{ijk}\bar{\Omega}_{jk},$$

whence, since $(\bar{g})_C=0$ by (2.2),

$$(2.8)\qquad \bar{\omega}^i = \frac{1}{2}c^{ijk}\left(\frac{\partial \bar{g}_{0k}}{\partial x^j}\right)_{\bar{C}}.$$

The differentiation with respect to $(\delta x^i)_{\bar{C}}$ has been performed and we now set $(\delta x^0)^-=0$. This does not alter (2.8), which thus expresses *the rotation vector* $\bar{\omega}^i$ *of the unstrained rod* in terms of $\bar{g}_{ij}$, as required.

3. **The strained rod.** When the rod defined in §2 is acted upon by certain external forces, it is displaced finitely and experiences a small strain. The

(7) For an explanation of the permutation symbols and their tensor character, see, for example, O. Veblen, *Invariants of Quadratic Differential Forms*, Cambridge, 1927, pp. 3–4, 25–26.

strained positions of the line of centroids $\overline{M}$, the cross-section $\overline{N}$, and the point $\overline{C}$ on both of them will be denoted by M, N and C respectively: M, N and C will be referred to as *the strained line of centroids, the strained cross-section and the strained centroid of the cross-section of the rod*, respectively. It will be noted that in general N is a curved surface.

In the unstrained rod, the coordinates associated with a general particle are x^i $(i=0, 1, 2)$. In the strained rod the same coordinates will be used to specify the same particle. Such coordinates are called co-moving coordinates. They are a particular set belonging to the more general coordinates used by Murnaghan(8). Thus, the line element in the strained rod will be of the form

$$ds^2 = g_{ij}dx^i dx^j, \tag{3.1}$$

where g_{ij} is the metric tensor and $g_{ij}=g_{ji}$.

In §2, a rotation vector $\bar{\omega}^i$ was determined for the unstrained rod. For the strained rod, a rotation vector ω^i can be defined somewhat analogously: ω^0 and ω^α will be referred to respectively as the twist and curvature of the strained rod. We shall now derive (3.7), which expresses ω^i in terms of g_{ij} and is thus the analogue of (2.8).

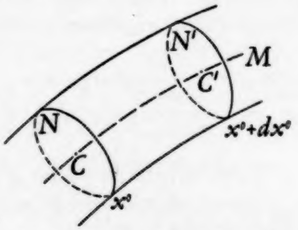


FIG. 3. A part of the strained rod

Two adjacent points C (x^0) and C' (x^0+dx^0) are taken on the line of centroids M (Figure 3): N and N' denote the cross-sections at C and C', respectively. The positions relative to C of points lying in N and near C will be denoted by infinitesimal vectors $(\delta x^\alpha)_C$. Similarly the positions relative to C' of points lying in N' will be denoted by $(\delta x^\alpha)_{C'}$. Corresponding to any point in N and near C there is a point in N' for which $(\delta x^\alpha)_C=(\delta x^\alpha)_{C'}$, and each point in N can be brought into coincidence with the corresponding point in N' by certain displacements. These displacements can be accomplished in two steps: (i) a rigid-body translation in which C moves into coincidence with C' and N moves into a new position N''; (ii) a system of displacements in which each point in N'' near C' moves into coincidence with the corresponding point in N'.

In step (i), each vector $(\delta x^\alpha)_C$ is propagated parallelly from C to C'. Just

(8) F. D. Murnaghan, American Journal of Mathematics, vol. 59 (1937), pp. 235–260.

as in §2, it is necessary to assume that the infinitesimal vectors $(\delta x^\alpha)_C$ have a third component $(\delta x^0)_C$ which can be allowed to vanish only after ω^i has been evaluated. Thus $(\delta x^i)_C$ becomes an infinitesimal vector in N'' at C' with components

$$\delta x^i - (F^i_{j0})_C(\delta x^j)_C dx^0, \tag{3.2}$$

where F^i_{jk} is the Christoffel symbol of the second kind in g_{ij},

$$F^i_{jk} = \tfrac{1}{2}g^{il}\left(\frac{\partial g_{kl}}{\partial x^j} + \frac{\partial g_{lj}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^l}\right): \tag{3.3}$$

g^{ij} is the minor of g_{ij} in the determinant $|g_{ij}|$ divided by that determinant, and $(F^i_{jk})_C$ represents F^i_{jk} evaluated at C (where $x^\alpha = 0$).

As a result of step (ii), (3.2) becomes $(\delta x^i)_{C'}$; and since $(\delta x^i)_C = (\delta x^i)_{C'}$, the system of infinitesimal displacements in step (ii) is

$$u^i = (F^i_{j0})_C(\delta x^j)_C dx^0. \tag{3.4}$$

This system of displacements can be considered as accomplished by means of a hypothetical homogeneous strain. This strain can be resolved into a pure strain and a rigid-body rotation. The rigid-body rotation is specified by the skew-symmetric tensor

$$\Omega_{ij} = \frac{1}{2}\left(\frac{\partial u_j}{\partial(\delta x^i)_C} - \frac{\partial u_i}{\partial(\delta x^j)_C}\right) = \frac{1}{2}dx^0\left(\frac{\partial g_{0j}}{\partial x^i} - \frac{\partial g_{0i}}{\partial x^j}\right). \tag{3.5}$$

It can also be specified by a contravariant vector $\omega^i dx^0$ such that $\lambda_i\omega^i dx^0$ is the angle of rotation about a direction specified by a unit vector λ_i. If $g = |g_{ij}|$ and c^{ijk} are the permutation symbols, then

$$\omega^i dx^0 = \tfrac{1}{2}(g^{-1/2})_C c^{ijk}\Omega_{jk}, \tag{3.6}$$

whence

$$\omega^i = \tfrac{1}{2}c^{ijk}\left(g^{-1/2}\frac{\partial g_{0k}}{\partial x^j}\right)_C. \tag{3.7}$$

The differentiation with respect to $(\delta x^i)_C$ has been performed and we now set $(\delta x^0)_C = 0$. This does not alter (3.7), which thus expresses *the rotation vector* ω^i *of the strained rod* in terms of g_{ij}, as required. This rotation vector refers to the rotation, as C moves along M, of the principal axes at C of the hypothetical strain introduced above. It will be noted that it does not refer to the rotation of a triad of vectors associated with the coordinate axes at C as C moves along the line of centroids M.

The elongation per unit length e of an element in the rod is accurately

$$e = \frac{ds - d\bar{s}}{d\bar{s}} = (g_{ij}dx^i dx^j)^{1/2}(\bar{g}_{kl}dx^k dx^l)^{-1/2} - 1. \tag{3.8}$$

For the line of centroids,

$$dx^{\alpha} = 0,$$

$\bar{g}_{ij}$ are as given in (2.2), and thus

$$e = (g_{00})_C^{1/2} - 1, \tag{3.9}$$

where the subscript C means that the expression in parenthesis is to be evaluated at C on the line of centroids.

It will be observed that all arguments in §3 are purely geometrical, and do not actually involve any assumption that the strain is small.

4. **Stress-strain relations.** The strain tensor(9) e_{ij} is defined by

$$2e_{ij} = g_{ij} - \bar{g}_{ij}. \tag{4.1}$$

Since the strain is to be small, the elongation per unit length of each element in the rod must be much smaller than unity. Thus, by (3.8), e_{ij} must be $\ll 1$.

The usual stress-strain relations for an isotropic elastic body involve two elastic constants, and in their development all powers of the stress and strain tensors above the first are neglected. Stress-strain relations which involve five elastic constants, and in the development of which all powers of the stress and strain tensors above the second are neglected, have been deduced from considerations of elastic energy by Murnaghan(10). In this paper we shall require the latter relations, and in terms of a *reduced* stress tensor T_{ij} defined by

$$E_{ij} = ET_{ij}, \tag{4.2}$$

E being Young's modulus and E_{ij} the stress tensor, they take the form

$$\begin{aligned} e_{ij} = &- \sigma K_1 g_{ij} + (1+\sigma)T_{ij} - (\sigma_1 (K_1)^2 + \sigma_2 K_2)g_{ij} \\ &+ (1 + \sigma + \sigma_2)K_1 T_{ij} + 2(1+\sigma)^2 T_{ik} g^{kl} T_{lj} + \sigma_3 K_{ij}, \end{aligned} \tag{4.3}$$

where σ is Poisson's ratio, $\sigma_1, \sigma_2, \sigma_3$ are three new elastic constants which are dimensionless, and

$$\begin{aligned} K_1 &= T^{i}_{\cdot i}, \\ 2K_2 &= \delta^{ij}_{kl} T^{k}_{i} T^{l}_{j}, \\ 2K_{ij} &= g_{ip} \delta^{pkl}_{jmn} T^{m}_{k} T^{n}_{l}, \end{aligned} \tag{4.4}$$

δ^{ij}_{kl}, δ^{pkl}_{jmn} being Kronecker deltas(11). If on the right side of (4.3) only the first two terms are retained, the two-constant stress-strain relations result.

---

(9) For a more detailed account of the application of tensor notation to elasticity, see, for example, P. Appell, *Mécanique Rationnelle*, vol. 5, Paris, 1926, p. 91.

(10) F. D. Murnaghan, loc. cit. See also L. Brillouin, Annales de Physique, vol. 3 (1925), pp. 251–297.

(11) For an explanation of the Kronecker deltas, see, for example, O. Veblen, loc. cit.

5. **Equations of equilibrium and compatibility.** Except in §2, which dealt with the geometry of the unstrained rod, throughout this paper the metric tensor g_{ij} of the strained rod and its tensor reciprocal g^{ij} will be used for lowering and raising suffixes, and also for forming tensorial derivatives.

Body forces are neglected. Thus, in terms of the reduced stress tensor T_{ij}, *the equations of equilibrium* are

$$T^{ij}{}_{|j} = 0, \tag{5.1}$$

where $|j$ signifies the covariant derivative. Since the space defined for the strained rod by (3.1) is flat, its curvature tensor R_{ijkl} must vanish, i.e.,

$$\begin{aligned} 2R_{ijkl} = \frac{\partial^2 g_{il}}{\partial x^j \partial x^k} + \frac{\partial^2 g_{jk}}{\partial x^i \partial x^l} - \frac{\partial^2 g_{ik}}{\partial x^j \partial x^l} - \frac{\partial^2 g_{jl}}{\partial x^i \partial x^k} \\ + 2g_{np}(\overset{n}{F}_{jk}\overset{p}{F}_{il} - \overset{n}{F}_{jl}\overset{p}{F}_{ik}) = 0, \end{aligned} \tag{5.2}$$

where $\overset{i}{F}_{jk}$ are defined in (3.3). Contained in (5.2) are just six distinct equations,

$$\begin{aligned} R_{0101} = R_{0102} = R_{0202} = 0, \\ R_{1220} = R_{1201} = R_{1212} = 0. \end{aligned} \tag{5.3}$$

In elastic problems involving small displacement only, (5.3) reduce to the usual equations of compatibility. Because of this, in the present work (5.3) will be called *the equations of compatibility*.

We let n_i denote the covariant components of a unit vector pointing along the outward normal to the lateral surface of the rod. Thus, since no stress is transmitted across this surface,

$$T_{ij}g^{ik}n_k = 0. \tag{5.4}$$

6. **Macroscopic equations of equilibrium.** In this paper it is found necessary to introduce certain equations which express the equilibrium of each element of the rod lying between adjacent cross-sections. These equations will be called the *macroscopic equations of equilibrium*, the word "macroscopic" indicating that the equations involve an integration over the cross-section of the rod.

In Figure 3, C is a point on the line of centroids M of the strained rod and N is the cross-section containing C. The reaction across N is statically equivalent to a force at C and a couple. The force will be denoted by EP_i and the couple by EQ_i, where E is Young's modulus. Thus, if $(\lambda^i)_C$ is any unit vector at C, then $EP_i(\lambda^i)_C$ and $EQ_i(\lambda^i)_C$ are, respectively, the component of the force in the direction $(\lambda^i)_C$ and the moment of the couple about that direction. The quantities P_i and Q_i are functions of x^0 only. They will be called *reduced force* and *reduced couple*, and are actually a force and a couple divided by Young's modulus. They satisfy the relations

$$P_i(\lambda^i)_C = \int_N T_{ij}\nu^i\lambda^j dS, \tag{6.1}$$

$$Q_i(\lambda^i)_C = \int_N g^{-1/2}c^{ijk}\lambda_i s\mu_j T_{kl}\nu^l dS, \tag{6.2}$$

where dS is an element of area on N, T_{ij} is the reduced stress tensor, c^{ijk} is a permutation symbol, g is the determinant $|g_{ij}|$, λ^i is a unit vector defined at a point on N by parallel propagation of $(\lambda^i)_C$ from C, ν^i is a unit vector normal to N in the direction of increasing x^0, μ_i is a unit vector at x^α on N tangent to the straight line drawn from C, and s is the length of this straight line. The explicit evaluation of P_i and Q_i will be accomplished later (§8) by use of (6.1) and (6.2); and in preparation for this, expressions for λ^i, ν^i and $s\mu_i$ will now be derived in turn.

At points on N

$$\lambda^i = (\lambda^i)_C + x^\alpha\left(\frac{\partial\lambda^i}{\partial x^\alpha}\right)_C + \cdots, \tag{6.3}$$

and since λ^i is propagated parallelly,

$$\frac{\partial\lambda^i}{\partial x^j} + F^i_{jk}\lambda^k = 0. \tag{6.4}$$

Thus

$$\lambda^i = (\lambda^i)_C - x^\alpha(F^i_{\alpha k})_C(\lambda^k)_C + \cdots. \tag{6.5}$$

For ν^i we have

$$\nu_0 = (g^{00})^{-1/2}, \quad \nu_\alpha = 0, \quad \nu^0 = (g^{00})^{1/2}, \quad \nu^\alpha = g^{\alpha 0}(g^{00})^{-1/2}. \tag{6.6}$$

The evaluation of $s\mu_i$ is more difficult. In order to give full power to the indicial notation, the coordinates of the initial point of the straight line of which s is the length and μ_i the unit tangent vector at x^i will be denoted by a^i. A straight line is a geodesic in three dimensions, and from variation of the end point of such a geodesic, it is known that

$$\mu_i = \frac{\partial s}{\partial x^i}. \tag{6.7}$$

Also, for such a geodesic as long as the point x^i is near the point a^i, it can be shown that

$$\begin{aligned} s^2 = {} & (g_{ij})_{x^l=a^l}(x^i - a^i)(x^j - a^j) \\ & + (g_{ij}F^j_{mn})_{x^l=a^l}(x^i - a^i)(x^m - a^m)(x^n - a^n) + \cdots, \end{aligned} \tag{6.8}$$

correct to the third order in $x^i - a^i$. Thus by (6.7),

$$2s\mu_i = \frac{\partial s^2}{\partial x^i} = 2(g_{ij})_{x^l=a^l}(x^j - a^j)$$
$$(6.9) \qquad + (g_{ij}F^j_{mn})_{x^l=a^l}(x^m - a^m)(x^n - a^n)$$
$$+ 2(g_{mn}F^n_{ij})_{x^l=a^l}(x^m - a^m)(x^j - a^j) + \cdots .$$

In order to evaluate $s\mu_i$ as it occurs in (6.2), we set

$$a^0 = x^0, \qquad a^\alpha = 0,$$

obtaining

$$(6.10) \qquad s\mu_i = (g_{i\alpha})_C x^\alpha + \tfrac{1}{2}(g_{ij}F^j_{\alpha\beta})_C x^\alpha x^\beta + (g_{\alpha n}F^n_{i\beta})_C x^\alpha x^\beta + \cdots .$$

In Figure 3, it is recalled that C (x^0) and C' (x^0+dx^0) are adjacent points on the line of centroids M of the strained rod, and that N and N' are the cross-sections containing C and C', respectively. The conditions of statical equilibrium for the part of the rod lying between N and N' are

$$(6.11) \qquad -(\lambda^i P_i)_C + (\lambda^i P_i)_{C'} = 0,$$

$$(6.12) \qquad -(\lambda^i Q_i)_C + (\lambda^i Q_i)_{C'} + c^{ijk}(g^{-1/2}\lambda_i s\mu_j P_k)_{C'} = 0,$$

where P_i, Q_i and g are as defined above, $(\lambda^i)_{C'}$ is a unit vector at C' produced by parallel propagation of $(\lambda^i)_C$ from C, μ_i is the unit vector at C' tangent to the geodesic drawn from C and s is the length of this geodesic. Now

$$(P_i)_{C'} = (P_i)_C + \left(\frac{dP_i}{dx^0}\right)_C dx^0 + \cdots ,$$
$$(6.13)$$
$$(Q_i)_{C'} = (Q_i)_C + \left(\frac{dQ_i}{dx^0}\right)_C dx^0 + \cdots ,$$

$$(6.14) \qquad (\lambda^i)_{C'} = (\lambda^i)_C + \left(\frac{\partial\lambda^i}{\partial x^0}\right)_C dx^0 + \cdots .$$

Since λ^i is propagated parallelly along M, (6.14) can be written in the form

$$(6.15) \qquad (\lambda^i)_{C'} = (\lambda^i)_C - (F^i_{j0})_C(\lambda^j)_C dx^0 + \cdots ,$$

where F^i_{jk} is the Christoffel symbol defined in (3.3). To evaluate $s\mu_i$ as it occurs in (6.12), we set $x^\alpha = a^\alpha = 0$, $x^0 - a^0 = dx^0$ in (6.10). Then by substitution for $s\mu_i$, and for $(\lambda^i)_{C'}$ from (6.15), it is found that (6.11) and (6.12) take the following forms, correct to the first order in dx^0,

$$(6.16) \qquad \left(\lambda^i \frac{dP_i}{dx^0} - F^j_{i0}P_j\lambda^i\right)_C = 0,$$

$$(6.17) \qquad \left(\lambda^i \frac{dQ_i}{dx^0} - F^j_{i0}Q_j\lambda^i + g^{-1/2}c^{jkl}g_{j0}P_k g_{li}\lambda^i\right)_C = 0,$$

whence, since $(\lambda^i)_C$ is an arbitrary unit vector,

$$\frac{dP_i}{dx^0} - (F^j_{i0})_C P_j = 0, \tag{6.18}$$

$$\frac{dQ_i}{dx^0} - (F^j_{i0})_C Q_j + (g^{-1/2} g_{ij} c^{jkl} g_{k0})_C P_l = 0. \tag{6.19}$$

These are the macroscopic equations of equilibrium. They can be found easily in vector form by vector methods. However, in that form they do not lend themselves to the method of approximation used in the present paper.

7. **The method of approximation.** The metric tensor g_{ij} of the strained rod, the strain tensor e_{ij} and the reduced stress tensor T_{ij} are all unknown. They are interrelated by (4.1) and by the stress-strain relations (4.3). They are to be determined by solving the equations of equilibrium (5.1) and the equations of compatibility (5.3). In order to solve these equations, approximation is introduced based on the thinness of the rod. In previous treatments of the problem, the manner in which this approximation was introduced suffered from certain defects which have been mentioned earlier. In this paper the approximation is introduced in the following systematic way.

We consider a singly infinite sequence of thin rods with identical lines of centroids of equal finite lengths, and similar cross-sections with average diameters continuously decreasing to zero. The relation

$$\epsilon = \frac{\textit{average diameter of cross-section of unstrained rod}}{\textit{length of unstrained rod}} \tag{7.1}$$

thus defines a dimensionless small parameter ϵ ranging continuously toward zero over values smaller than unity.

With regard to order of magnitude, a quantity q will be referred to as of order ϵ^r ($r=0, 1, 2, \cdots$) if q/ϵ^r tends to a non-infinite, nonzero limit as ϵ tends to zero. A quantity of order ϵ^0 will be referred to as finite. For example, if l is the length of the line of centroids of the unstrained rod and E is Young's modulus, then $l\epsilon^r$ is a length of order ϵ^r, $El^2\epsilon^r$ and $El^3\epsilon^r$ are respectively a force and a couple of order ϵ^r, ϵ^r/l is a curvature of order ϵ^r, and so on for reduced force, reduced couple, etc.

We limit ourselves to rods of uniform cross-section, and since all rods in the sequence have similar cross-sections, the equation of their lateral surfaces takes the form

$$f\left(\frac{x^1}{\epsilon}, \frac{x^2}{\epsilon}\right) = 0. \tag{7.2}$$

It will be recalled that x^α ($\alpha=1, 2$) are rectangular cartesian coordinates in a general normal cross-section of the unstrained rod, while x^0 is the arc-length

of the line of centroids of the unstrained rod. Since the range of x^α is much smaller than that of x^0, new coordinates are introduced to replace x^α. They are defined by

$$x^\alpha = \epsilon\xi^\alpha, \tag{7.3}$$

so that the ranges of x^0 and ξ^α are all finite. Thus (7.2) becomes

$$f(\xi^1, \xi^2) = 0. \tag{7.4}$$

The metric tensor $\bar{g}_{ij}$ of the unstrained rod is assumed known. It is a function of x^0 and the relatively small coordinates x^α. It is assumed that $\bar{g}_{ij}$ can be expanded in Taylor series in x^α, and since

$$(\bar{g}_{ij})_{x^\gamma=0} = \delta^i_j$$

(by (2.2)), we then have

$$\bar{g}_{ij} = \delta^i_j + \epsilon\xi^\alpha \bar{g}_{ij}^{(\alpha)} + \tfrac{1}{2}\epsilon^2\xi^\alpha\xi^\beta \bar{g}_{ij}^{(\alpha\beta)} + \cdots, \tag{7.5}$$

where

$$\bar{g}_{ij}^{(\alpha)} = \left(\frac{\partial \bar{g}_{ij}}{\partial x^\alpha}\right)_{x^\gamma=0}, \qquad \bar{g}_{ij}^{(\alpha\beta)} = \left(\frac{\partial^2 \bar{g}_{ij}}{\partial x^\alpha \partial x^\beta}\right)_{x^\gamma=0}, \cdots. \tag{7.6}$$

There are similar expansions for $\bar{b}^{ij}$,

$$\bar{b}^{ij} = \delta^i_j + \epsilon\xi^\alpha \bar{b}^{ij(\alpha)} + \tfrac{1}{2}\epsilon^2\xi^\alpha\xi^\beta \bar{b}^{ij(\alpha\beta)} + \cdots, \tag{7.7}$$

where the coefficients $\bar{b}^{ij(\alpha)}$, $\bar{b}^{ij(\alpha\beta)}$, $\cdots$ are defined by equations analogous to (7.6), and since the relations $\bar{g}_{jk}\bar{b}^{ik} = \delta^i_j$ are true for all values of ϵ and ξ^α, these coefficients are related to the coefficients in (7.5) by

$$\bar{b}^{ij(\alpha)} = -\bar{g}_{ij}^{(\alpha)}, \qquad \bar{b}^{ij(\alpha\beta)} = 2\bar{g}_{ik}^{(\alpha)}\bar{g}_{jk}^{(\beta)} - \bar{g}_{ij}^{(\alpha\beta)}, \cdots. \tag{7.8}$$

We note that $\bar{g}_{ij}$ is finite when $i=j$, and is of order ϵ when $i \neq j$.

It will be assumed that the components of the unknown metric tensor g_{ij} of the strained rod in equilibrium admit expansions in powers of ϵ,

$$g_{ij} = g_{ij(0)} + \epsilon g_{ij(1)} + \epsilon^2 g_{ij(2)} + \cdots, \tag{7.9}$$

the coefficients being regular functions of x^0, ξ^α. From (4.1) it then follows that there are expansions for the strain tensor

$$e_{ij} = e_{ij(0)} + \epsilon e_{ij(1)} + \cdots, \tag{7.10}$$

and because of the stress-strain relations, for the reduced stress tensor,

$$T_{ij} = T_{ij(0)} + \epsilon T_{ij(1)} + \cdots. \tag{7.11}$$

To confine our attention to small strain, we must limit ourselves to those cases where

$$e_{ij(0)} = 0. \tag{7.12}$$

Then, since strain and stress must vanish simultaneously to the same order of magnitude,

$$T_{ij(0)} = 0. \tag{7.13}$$

When we substitute in (4.1) the power series (7.5), (7.9) and (7.10) and equate coefficients of powers of ϵ, we obtain, by use of (7.12),

$$g_{ij(0)} = \delta^i_j,\quad 2e_{ij(1)} = g_{ij(1)} - \bar{g}_{ij}^{(\alpha)}\xi^\alpha,\quad 2e_{ij(2)} = g_{ij(2)} - \tfrac{1}{2}\bar{g}_{ij}^{(\alpha\beta)}\xi^\alpha\xi^\beta, \cdots . \tag{7.14}$$

These equations relate the coefficients in (7.9) and (7.10). By means of the stress-strain relations (4.3), we can also relate the coefficients in (7.10) and (7.11), or (7.9) and (7.11). Only the latter relations will be required. They appear in the following section (equations (8.10), (8.11)).

In (7.9), (7.10) and (7.11), the coefficients of ϵ^0 are known. If for any problem we determine all the remaining coefficients, we shall have solved that problem exactly. If we determine only the coefficients up to and including those of ϵ^n, we shall have obtained an approximate solution of the problem to within an error of order ϵ^{n+1}, and the smaller ϵ (i.e., the thinner the rod), the better the approximation. In order to derive equations to be solved for the unknown coefficients in (7.9), (7.10) and (7.11), we shall express as power series in ϵ with coefficients independent of ϵ the left sides of the equations of equilibrium (5.1), the equations of compatibility (5.3) and the macroscopic equations of equilibrium (6.18) and (6.19), and then equate to zero the coefficients of the various powers of ϵ. To this end, we must express as power series in ϵ the quantities g^{ij}, F^i_{jk}, n_i, P_i, Q_i, $\bar{\omega}^i$, ω^i and e. This will occupy the remainder of this section.

In view of (7.9), g^{ij} can be expressed in the form

$$g^{ij} = \delta^i_j + \epsilon g^{ij}_{(1)} + \epsilon^2 g^{ij}_{(2)} + \cdots , \tag{7.15}$$

and since the relations

$$g_{jk}g^{ik} = \delta^i_j$$

are true for all values of ϵ, it is found that

$$g^{ij}_{(1)} = -g_{ij(1)},\qquad g^{ij}_{(2)} = g_{ik(1)}g_{jk(1)} - g_{ij(2)}, \cdots . \tag{7.16}$$

Differentiation with respect to x^0 and ξ^α will be denoted by a subscript preceded by a comma. Thus

$$\phi_{,0} = \frac{\partial\phi}{\partial x^0},\qquad \phi_{,\alpha} = \frac{\partial\phi}{\partial\xi^\alpha}. \tag{7.17}$$

The Christoffel symbols of the second kind Γ^i_{jk} are functions of g^{ij} and the derivatives of g_{ij} with respect to x^i. By substitution, they can thus be expressed in the form

$$\Gamma^i_{jk} = \Gamma^i_{jk(0)} + \epsilon\Gamma^i_{jk(1)} + \cdots, \tag{7.18}$$

where

$$\begin{aligned} &2\Gamma^0_{00(0)} = 0, \qquad 2\Gamma^\alpha_{00(0)} = -g_{00(1),\alpha}, \qquad 2\Gamma^0_{0\alpha(0)} = g_{00(1),\alpha},\\ &2\Gamma^\beta_{\alpha0(0)} = g_{\beta0(1),\alpha} - g_{0\alpha(1),\beta}, \qquad 2\Gamma^0_{\alpha\beta(0)} = g_{\beta0(1),\alpha} + g_{0\alpha(1),\beta},\\ &2\Gamma^\gamma_{\alpha\beta(0)} = g_{\beta\gamma(1),\alpha} + g_{\gamma\alpha(1),\beta} - g_{\alpha\beta(1),\gamma}, \end{aligned} \tag{7.19}$$

$$\begin{aligned} &2\Gamma^0_{00(1)} = g_{00(1),0} + g_{0\alpha(1)}g_{00(1),\alpha},\\ &2\Gamma^0_{0\alpha(1)} = g_{00(2),\alpha} - g_{00(1)}g_{00(1),\alpha} - g_{0\beta(1)}(g_{\beta0(1),\alpha} - g_{0\alpha(1),\beta}). \end{aligned} \tag{7.20}$$

There are similar expressions for the remaining components of $\Gamma^i_{jk(1)}$, and for $\Gamma^i_{jk(2)}$, $\Gamma^i_{jk(3)}$, $\cdots$, which will not be required in the present investigation.

We shall now express as a power series in ϵ the unit covariant vector n_i pointing along the outward normal to the lateral surface of the strained rod. The equation of the lateral surface of the rod is (7.4). Thus

$$n_0 = 0, \qquad n_\alpha = kf_{,\alpha}, \tag{7.21}$$

where $k^{-2}=g^{\alpha\beta}f_{,\alpha}f_{,\beta}$. In (7.15), $g^{\alpha\beta}$ is expressed as a power series in ϵ. Thus k and hence also n_i can be expressed as a power series in ϵ. We obtain

$$n_0 = 0, \qquad n_\alpha = n_{\alpha(0)} + \epsilon n_{\alpha(1)} + \cdots, \tag{7.22}$$

where

$$n_{\alpha(0)} = f_{,\alpha}(f_{,\beta}f_{,\beta})^{-1/2}, \quad n_{\alpha(1)} = \tfrac{1}{2}f_{,\alpha}(f_{,\beta}f_{,\beta})^{-3/2}g_{\gamma\delta(1)}f_{,\gamma}f_{,\delta}, \cdots. \tag{7.23}$$

We note that if $d\xi^\alpha$ are the components of an arbitrary displacement on the periphery of a cross-section N, then $f_{,\alpha}d\xi^\alpha=0$, whence $f_{,1}/d\xi^2=-f_{,2}/d\xi^1$. Thus (7.23) can be written in the form

$$n_{1(0)} = \frac{d\xi^2}{d\tau}, \qquad n_{2(0)} = -\frac{d\xi^1}{d\tau}, \cdots, \tag{7.24}$$

where

$$(d\tau)^2 = d\xi^\alpha d\xi^\alpha.$$

The reaction across a cross-section of the strained rod is specified by a certain reduced force P_i and a reduced couple Q_i satisfying (6.1) and (6.2), respectively. An explicit evaluation of P_i and Q_i as power series in ϵ will now be made. The vectors λ^i, ν^i and $s\mu_i$ in (6.1) and (6.2) are given by (6.5), (6.6) and (6.10). They can thus be expressed as power series in ϵ by substituting for x^α, g_{ij}, g^{ij}, Γ^i_{jk} from (7.3), (7.9), (7.15) and (7.18). We then obtain

$$\lambda^i = (\lambda^i)_C - \epsilon\xi^\alpha (F^i_{\alpha k})_C (\lambda^k)_C + O(\epsilon^2), \tag{7.25}$$

$$\nu^0 = 1 - \tfrac{1}{2}\epsilon g_{00(1)} + O(\epsilon^2), \qquad \nu^\alpha = -\epsilon g_{\alpha 0(1)} + O(\epsilon^2), \tag{7.26}$$

$$\begin{aligned} s\mu_0 &= \epsilon^2\left[(g_{0\alpha(1)})_C\xi^\alpha + \tfrac{1}{2}(g_{0\alpha(1),\beta})_C\xi^\alpha\xi^\beta\right] + O(\epsilon^3), \\ s\mu_\alpha &= \epsilon\xi^\alpha + \epsilon^2\left[(g_{\alpha\beta(1)})_C\xi^\beta + \tfrac{1}{4}(2g_{\alpha\beta(1),\gamma} + g_{\beta\gamma(1),\alpha})_C\xi^\beta\xi^\gamma\right] + O(\epsilon^3), \end{aligned} \tag{7.27}$$

where $O(\epsilon^r)$ means terms of order ϵ^r and higher. The element of area dS on N is given by

$$dS = (g')^{1/2}dx^1dx^2 = \epsilon^2(g')^{1/2}d\xi^1d\xi^2, \tag{7.28}$$

where

$$g' = |g_{\alpha\beta}| = 1 + \epsilon g_{\alpha\alpha(1)} + O(\epsilon^2). \tag{7.29}$$

Of the quantities occurring on the right-hand sides of (6.1) and (6.2), T_{ij}, λ^i, ν^i, $s\mu_i$ and dS are expressed in (7.11), (7.25), (7.26), (7.27) and (7.28) as power series in ϵ; by (7.9), $g=|g_{ij}|$ can be similarly expressed, and thus the entire right-hand sides of (6.1) and (6.2) can be expressed as power series in ϵ. In the resulting equations, $(\lambda^i)_C$ appears in every term, and since $(\lambda^i)_C$ is an arbitrary unit vector at C, we conclude that

$$P_i = \epsilon^3 P_{i(3)} + \epsilon^4 P_{i(4)} + \cdots, \tag{7.30}$$

$$Q_i = \epsilon^4 Q_{i(4)} + \epsilon^5 Q_{i(5)} + \cdots, \tag{7.31}$$

where

$$P_{i(3)} = \int_N T_{0i(1)}d\xi^1d\xi^2, \tag{7.32}$$

$$\begin{aligned} P_{i(4)} = \int_N \big[&T_{0i(2)} + \tfrac{1}{2}(g_{\alpha\alpha(1)} - g_{00(1)})T_{0i(1)} \\ &- g_{\alpha 0(1)}T_{\alpha i(1)} - T_{0j(1)}\xi^\alpha(F^j_{\alpha i(0)})_C\big]d\xi^1d\xi^2, \end{aligned} \tag{7.33}$$

$$Q_{i(4)} = \int_N c^{i\alpha k}\xi^\alpha T_{k0(1)}d\xi^1d\xi^2, \tag{7.34}$$

$$\begin{aligned} Q_{i(5)} = \int_N \big[&c^{j\alpha k} g_{ji(1)}\xi^\alpha T_{k0(1)} - c^{j\alpha k}(F^j_{\beta i(0)})_C\xi^\beta\xi^\alpha T_{k0(1)} \\ &+ c^{i\alpha k}\left\{(g_{\alpha\beta(1)})_C\xi^\beta + \tfrac{1}{4}(2g_{\alpha\beta(1),\gamma} + g_{\beta\gamma(1),\alpha})_C\xi^\beta\xi^\gamma\right\}T_{k0(1)} \\ &+ c^{i\alpha k}\xi^\alpha\left\{T_{k0(2)} - T_{kj(1)}g_{j0(1)}\right\} \\ &+ c^{i0k}\left\{(g_{0\alpha(1)})_C\xi^\alpha + \tfrac{1}{2}(g_{0\alpha(1),\beta})_C\xi^\alpha\xi^\beta\right\}T_{k0(1)}\big]d\xi^1d\xi^2. \end{aligned} \tag{7.35}$$

By (7.5), $\bar{g}=|\bar{g}_{ij}|$ can be expressed as a power series in ϵ. Thus, by (2.8) we are able to express the rotation vector $\bar{\omega}^i$ of the unstrained rod in the form

$$\bar{\omega}^{i} = \tfrac{1}{2}c^{i\alpha k}\bar{g}_{0k}^{(\alpha)}. \tag{7.36}$$

By (7.9), $g=|g_{ij}|$ can also be expressed as a power series in ϵ. Thus, by (3.7) we can express the rotation vector ω^i of the strained rod in the form

$$\omega^{i} = \omega^{i}_{(0)} + \epsilon\omega^{i}_{(1)} + \cdots, \tag{7.37}$$

where

$$\omega^{i}_{(0)} = \tfrac{1}{2}c^{i\alpha k}(g_{0k(1),\alpha})_C, \tag{7.38}$$

$$\omega^{i}_{(1)} = \tfrac{1}{2}c^{i\beta k}(g_{0k(2),\beta} - \tfrac{1}{2}g_{jj(1)}g_{0k(1),\beta})_C + \tfrac{1}{2}c^{i0\beta}(g_{0\beta(1),0})_C. \tag{7.39}$$

By (3.9) we can also express e, the elongation per unit length of the line of centroids, in the form

$$e = \epsilon e_{(1)} + \epsilon^2 e_{(2)} + \cdots, \tag{7.40}$$

where

$$e_{(1)} = \tfrac{1}{2}(g_{00(1)})_C, \tag{7.41}$$

$$e_{(2)} = \tfrac{1}{2}(g_{00(2)})_C - \tfrac{1}{8}(g_{00(1)})^2_C. \tag{7.42}$$

8. **The fundamental equations.** In this section, there will be developed equations for the determination of the unknown coefficients in the expansions (7.9), (7.10), (7.11) for g_{ij}, e_{ij}, T_{ij}, respectively.

The equations of equilibrium (5.1) can be written in the form

$$g^{jk}\left(\frac{\partial T_{ik}}{\partial x^j} - F^{l}_{jk}T_{il} - F^{l}_{ij}T_{kl}\right) = 0. \tag{8.1}$$

All quantities in this equation are expressed in (7.11), (7.15) and (7.18) as power series in ϵ. Hence we can express the left-hand side of (8.1) as a power series in ϵ, and since (8.1) is true for all small values of ϵ, the coefficients of the various powers of ϵ must vanish identically. The coefficients of ϵ^0 vanish if

$$T_{i\alpha(1),\alpha} = 0. \tag{8.2}$$

These are *the first order equations of equilibrium*. It is found that the coefficients of ϵ vanish if

$$\begin{aligned} &T_{i\alpha(2),\alpha} + T_{i0(1),0} - g_{j\alpha(1)}T_{ij(1),\alpha} - g_{0\alpha(1),\alpha}T_{i0(1)} \\ &\quad - \tfrac{1}{2}(2g_{\gamma\alpha(1),\alpha} - g_{jj(1),\gamma})T_{i\gamma(1)} - \tfrac{1}{2}\delta^{\alpha}_{i}g_{jk(1),\alpha}T_{jk(1)} = 0. \end{aligned} \tag{8.3}$$

These are *the second order equations of equilibrium*. Similarly there are equations of equilibrium of the third order, fourth order, and so on.

All quantities in the equations of compatibility (5.3) have also been expressed as power series in ϵ. The left sides of the equations themselves can then be expressed as power series in ϵ, and the coefficients of the various

powers of ϵ must vanish identically. The coefficients of ϵ^{-1} give rise to *the first order equations of compatibility*

$$g_{00(1),\alpha\beta} = 0, \tag{8.4}$$

$$\begin{aligned} g_{01(1),22} - g_{02(1),12} &= 0, \\ g_{02(1),11} - g_{01(1),21} &= 0, \end{aligned} \tag{8.5}$$

$$2g_{12(1),12} - g_{11(1),22} - g_{22(1),11} = 0, \tag{8.6}$$

while the coefficients of ϵ^0 give rise to *the second order equations of compatibility*

$$\begin{aligned} g_{00(2),\alpha\beta} &- g_{0\alpha(1),0\beta} - g_{0\beta(1),0\alpha} - \tfrac{1}{2}g_{00(1),\alpha}g_{00(1),\beta} \\ &- \tfrac{1}{2}(g_{0\gamma(1),\alpha} - g_{0\alpha(1),\gamma})(g_{0\gamma(1),\beta} - g_{0\beta(1),\gamma}) \\ &- \tfrac{1}{2}g_{00(1),\gamma}(g_{\beta\gamma(1),\alpha} + g_{\gamma\alpha(1),\beta} - g_{\alpha\beta(1),\gamma}) = 0, \end{aligned} \tag{8.7}$$

$$\begin{aligned} g_{01(2),22} &- g_{02(2),12} + g_{22(1),01} - g_{12(1),02} + g_{00(1),1}g_{02(1),2} \\ &+ \tfrac{1}{2}g_{\alpha\alpha(1),2}(g_{02(1),1} - g_{01(1),2}) - \tfrac{1}{2}g_{00(1),2}(g_{02(1),1} + g_{01(1),2}) = 0, \\ g_{02(2),11} &- g_{01(2),21} + g_{11(1),02} - g_{21(1),01} + g_{00(1),2}g_{01(1),1} \\ &+ \tfrac{1}{2}g_{\alpha\alpha(1),1}(g_{01(1),2} - g_{02(1),1}) - \tfrac{1}{2}g_{00(1),1}(g_{01(1),2} + g_{02(1),1}) = 0, \end{aligned} \tag{8.8}$$

$$\begin{aligned} 2g_{12(2),12} &- g_{11(2),22} - g_{22(2),11} - 2g_{01(1),1}g_{02(1),2} \\ &+ \tfrac{1}{2}(g_{2\gamma(1),1} + g_{\gamma1(1),2} - g_{12(1),\gamma})(g_{2\gamma(1),1} + g_{\gamma1(1),2} - g_{12(1),\gamma}) \\ &+ \tfrac{1}{2}(g_{02(1),1} + g_{01(1),2})^2 - \tfrac{1}{2}(2g_{1\gamma(1)} - g_{11(1),\gamma})(2g_{2\gamma(1),2} - g_{22(1),\gamma}) = 0. \end{aligned} \tag{8.9}$$

Similarly we can write down equations of compatibility of the third order, the fourth order, and so on.

By replacing e_{ij} in the stress-strain relations (4.3) by $\frac{1}{2}(g_{ij} - \bar{g}_{ij})$ and substituting for $\bar{g}_{ij}$, g_{ij} and T_{ij} from (7.5), (7.9) and (7.11) with $T_{ij(0)} = 0$ as per (7.13), in the same way we deduce *stress-strain relations of the first order*,

$$g_{ij(1)} = \bar{g}_{ij}^{(\alpha)}\xi^{\alpha} - 2\sigma\delta^{i}_{j}T_{kk(1)} + 2(1+\sigma)T_{ij(1)}, \tag{8.10}$$

and *stress-strain relations of the second order*,

$$\begin{aligned} g_{ij(2)} = \tfrac{1}{2}\bar{g}_{ij}^{(\alpha\beta)}\xi^{\alpha}\xi^{\beta} &- 2\sigma\delta^{i}_{j}T_{kk(2)} + 2(1+\sigma)T_{ij(2)} \\ &+ \delta^{i}_{j}[2\sigma g_{kl(1)}T_{kl(1)} - (2\sigma_1 + \sigma_2 - \sigma_3)T_{kk(1)}T_{ll(1)} \\ &+ (\sigma_2 - \sigma_3)T_{kl(1)}T_{kl(1)}] \\ &- 2\sigma T_{kk(1)}g_{ij(1)} + 2(1 + \sigma + \sigma_2 - \sigma_3)T_{kk(1)}T_{ij(1)} \\ &+ 2[2(1+\sigma)^2 + \sigma_3]T_{ik(1)}T_{jk(1)}. \end{aligned} \tag{8.11}$$

In the stress-strain relations (4.3), quantities of order $(e_{ij})^3$ were neglected. If such quantities had been retained, it would be possible to write down stress-strain relations of the third order. Similarly stress-strain relations of higher orders could be found.

The boundary conditions on the lateral surface of the rod are (5.4). By substitution in (5.4) from (7.11), (7.13), (7.15) and (7.22), we can convert the left sides into power series in ϵ. It is found that the coefficients of ϵ vanish if

$$T_{i\beta(1)}n_{\beta(0)} = 0. \tag{8.12}$$

These are *the first order boundary conditions*. Similarly the coefficients of ϵ^2 vanish if

$$T_{i\beta(2)}n_{\beta(0)} - T_{ij(1)}g_{j\beta(1)}n_{\beta(0)} + T_{i\beta(1)}n_{\beta(1)} = 0. \tag{8.13}$$

These are *the second order boundary conditions*. In the same way there could be written down boundary conditions of the third order, of the fourth order, and so on.

All quantities in the macroscopic equations of equilibrium (6.18) and (6.19) have been expressed in §7 as power series in ϵ. By substitution, the left sides of these equations can then be expressed as power series in ϵ and the coefficients of the various powers of ϵ equated to zero. The lowest power of ϵ occurring is ϵ^3, and its coefficients vanish if

$$\frac{dP_{i(3)}}{dx^0} - (F^{j}_{i0(0)})cP_{j(3)} = 0, \tag{8.14}$$

$$P_{\alpha(3)} = 0. \tag{8.15}$$

These are *the first order macroscopic equations of equilibrium*. The coefficients of ϵ^4 vanish if

$$\frac{dP_{i(4)}}{dx^0} - (F^{j}_{i0(0)})cP_{j(4)} - (F^{0}_{0i(1)})cP_{0(3)} = 0, \tag{8.16}$$

$$\frac{dQ_{i(4)}}{dx^0} - (F^{j}_{i0(0)})cQ_{j(4)} + c^{i\alpha 0}(g_{\alpha 0(1)})cP_{0(3)} + c^{i0\alpha}P_{\alpha(4)} = 0, \tag{8.17}$$

and these are *the second order macroscopic equations of equilibrium*. The coefficients of ϵ^5 vanish if

$$\frac{dP_{i(5)}}{dx^0} - (F^{j}_{i0(0)})cP_{j(5)} - (F^{j}_{i0(1)})cP_{j(4)} - (F^{0}_{0i(2)})cP_{0(3)} = 0, \tag{8.18}$$

$$\begin{aligned} &\frac{dQ_{i(5)}}{dx^0} - (F^{j}_{i0(0)})cQ_{j(5)} - (F^{j}_{i0(1)})cQ_{j(4)} + c^{i0\alpha}P_{\alpha(5)} \\ &\quad + (g_{i\alpha(1)} - \tfrac{1}{2}\delta^{\alpha}_{i}g_{nn(1)})cc^{\alpha 0\beta}P_{\beta(4)} + c^{ikl}(g_{k0(1)})cP_{l(4)} \\ &\quad + (\delta^{\alpha}_{i}g_{\beta 0(2)} + g_{i\alpha(1)}g_{\beta 0(1)} - \tfrac{1}{2}\delta^{\alpha}_{i}g_{\beta 0(1)}g_{nn(1)})cc^{\alpha\beta 0}P_{0(3)} = 0, \end{aligned} \tag{8.19}$$

where the required components of $F^{i}_{jk(r)}$ are given in (7.19) and (7.20). These are *the third order macroscopic equations of equilibrium*. Similarly there can be written down macroscopic equations of equilibrium of higher orders.

The fundamental equations listed in the previous part of this section are to be solved in the following manner:

I. T_{ij} is the reduced stress tensor and g_{ij} the metric tensor of the strained rod. In (7.11) and (7.9), they are expressed as power series in ϵ, the first terms $T_{ij(0)}$ and $g_{ij(0)}$ being given by (7.13) and (7.14). If any problem in thin rods is to be considered to the first approximation, it is only necessary to determine $T_{ij(1)}$ and $g_{ij(1)}$, and this is done by solving the first order stress-strain relations (8.10) with the first order equations of equilibrium and compatibility (8.2), (8.4), (8.5) and (8.6), subject to the first order boundary conditions (8.12) over the sides of the rod. These solutions for $T_{ij(1)}$ and $g_{ij(1)}$ will contain arbitrary functions of x^0. By substitution for $T_{ij(1)}$ and $g_{ij(1)}$ in (7.32) and (7.34) it is then possible to express $P_{i(3)}$ and $Q_{i(4)}$ in terms of these unknown functions of x^0, which can then be determined by means of the first and second order macroscopic equations of equilibrium (8.14), (8.15), (8.16) and (8.17).

II. If any problem is to be considered to the second approximation, it is necessary to determine $T_{ij(2)}$ and $g_{ij(2)}$ in addition to $T_{ij(1)}$ and $g_{ij(1)}$ already found as above. This is done by means of the second order stress-strain relations (8.11) and the second order equations of equilibrium and compatibility (8.3), (8.7), (8.8) and (8.9), subject to the second order boundary conditions (8.13). These solutions for $T_{ij(2)}$ and $g_{ij(2)}$ will also contain arbitrary functions of x^0. By substitution for $T_{ij(2)}$ and $g_{ij(2)}$ in (7.33) and (7.35) it is possible to express $P_{i(4)}$ and $Q_{i(5)}$ in terms of these unknown functions of x^0, which can then be determined by means of the third order macroscopic equations of equilibrium, (8.18) and (8.19).

III. By continuing in the same way, we can consider any problem to a third approximation by determining $T_{ij(3)}$ and $g_{ij(3)}$, to a fourth approximation by determining $T_{ij(4)}$ and $g_{ij(4)}$, and so on.

9. **The first approximation: the stress.** As indicated in §8, I, just above, the first order equations of equilibrium

$$T_{i\alpha(1),\alpha} = 0, \tag{9.1}$$

and the first order equations of compatibility

$$g_{00(1),\alpha\beta} = 0, \tag{9.2}$$

$$\begin{aligned} g_{01(1),22} - g_{02(1),12} &= 0, \\ g_{02(1),11} - g_{01(1),21} &= 0, \end{aligned} \tag{9.3}$$

$$2g_{12(1),12} - g_{11(1),22} - g_{22(1),11} = 0, \tag{9.4}$$

with the first order boundary conditions

$$T_{i\beta(1)}n_{\beta(0)} = 0, \tag{9.5}$$

are to be solved for $T_{ij(1)}$ and $g_{ij(1)}$, which are themselves related by the first order stress-strain relations

$$g_{ij(1)} = \overset{(\alpha)}{\bar{g}}_{ij}\xi^{\alpha} - 2\sigma\delta^{i}_{j}T_{kk(1)} + 2(1+\sigma)T_{ij(1)}. \tag{9.6}$$

To solve equations of this nature, recourse is usually made to Saint-Venant's hypothesis, which states that these equations imply that $T_{\alpha\beta(1)} = 0$. It is not necessary to make this hypothesis here, as in Appendix A it is proved that, of necessity,

$$T_{\alpha\beta(1)} = 0. \tag{9.7}$$

This constitutes a proof of Saint-Venant's hypothesis in the case of thin rods displaced finitely.

From (9.2) it follows that

$$g_{00(1)} = (\overset{(\gamma)}{g}_{00} + A_{\gamma})\xi^{\gamma} + B, \tag{9.8}$$

where A_{γ} and B are arbitrary functions of x^0. When $i=j=0$, (9.6) then gives

$$T_{00(1)} = \tfrac{1}{2}(A_{\alpha}\xi^{\alpha} + B). \tag{9.9}$$

We now solve (9.1) and (9.3) for $T_{0\alpha(1)}$ and $g_{0\alpha(1)}$. From (9.3),

$$g_{01(1),2} - g_{02(1),1} = D', \tag{9.10}$$

where D' is an arbitrary function of x^0. By (9.6), this gives

$$T_{01(1),2} - T_{02(1),1} = -D/(1+\sigma), \tag{9.11}$$

where $D = -\frac{1}{2}(D' - \bar{g}^{(2)}_{01} + \bar{g}^{(1)}_{02})$. From (9.11) and (9.1) with $i=0$ it follows that

$$T_{01(1)} = \frac{D}{2(1+\sigma)}\chi_{,2}, \qquad T_{02(1)} = -\frac{D}{2(1+\sigma)}\chi_{,1}, \tag{9.12}$$

where χ is a function of ξ^{α} only, satisfying

$$\chi_{,\alpha\alpha} = -2. \tag{9.13}$$

By (7.24), the boundary condition (9.5) with $i=0$ is satisfied if

$$\chi = 0 \tag{9.14}$$

on the lateral surface of the rod.

From (2.2), $\bar{g}^{(\gamma)}_{\alpha\beta} = 0$, and thus the collected results are

$$T_{\alpha\beta(1)} = 0, \qquad T_{0\alpha(1)} = \frac{D}{2(1+\sigma)}c_{\alpha\beta}\chi_{,\beta}, \qquad T_{00(1)} = \tfrac{1}{2}(A_{\alpha}\xi^{\alpha} + B), \tag{9.15}$$

$$\begin{aligned} g_{\alpha\beta(1)} &= -\sigma\delta^{\alpha}_{\beta}(A_{\gamma}\xi^{\gamma} + B), \\ g_{0\alpha(1)} &= \overset{(\gamma)}{\bar{g}}_{0\alpha}\xi^{\gamma} + Dc_{\alpha\beta}\chi_{,\beta}, \qquad g_{00(1)} = \overset{(\gamma)}{\bar{g}}_{00}\xi^{\gamma} + A_{\gamma}\xi^{\gamma} + B, \end{aligned} \tag{9.16}$$

$c_{\alpha\beta}$ being a permutation symbol and A_{α}, B and D arbitrary functions of x^0.

By means of (9.16), the leading coefficients in the expressions (7.37), and (7.40) for ω^i and e can now be expressed in the form

$$
(9.17)\qquad \begin{aligned}
\overset{0}{\omega}_{(0)} &= \tfrac{1}{2}c_{\alpha\beta}\overset{(\alpha)}{\bar{g}_{0\beta}} + D,\\
\overset{\alpha}{\omega}_{(0)} &= \tfrac{1}{2}c_{\alpha\beta}(\overset{(\beta)}{\bar{g}_{00}} + A_\beta),
\end{aligned}
$$

$$
(9.18)\qquad \begin{aligned}
\overset{0}{\omega}_{(1)} &= \tfrac{1}{2}c_{\beta\gamma}(g_{0\gamma(2),\beta})c - \tfrac{1}{4}(1-2\sigma)B(2D + c_{\beta\gamma}\overset{(\beta)}{\bar{g}_{0\gamma}}),\\
\overset{\alpha}{\omega}_{(1)} &= \tfrac{1}{2}c_{\alpha\beta}(g_{00(2),\beta})c - \tfrac{1}{4}(1-2\sigma)Bc_{\alpha\beta}(\overset{(\beta)}{\bar{g}_{00}} + A_\beta) + \tfrac{1}{2}\frac{dD}{dx^0}(\chi_{,\alpha})c,
\end{aligned}
$$

$$
(9.19)\qquad e_{(1)} = \tfrac{1}{2}B.
$$

10. **The first approximation: the macroscopic equations of equilibrium.** The arbitrary functions of x^0 occurring in (9.15) and (9.16) are now to be determined in the manner outlined in §8, I. We find that

$$
(10.1)\qquad P_{0(3)} = \tfrac{1}{2}BA, \qquad P_{\alpha(3)} = 0,
$$

$$
(10.2)\qquad Q_{0(4)} = DT, \qquad Q_{1(4)} = \tfrac{1}{2}A_2I_1, \qquad Q_{2(4)} = -\tfrac{1}{2}A_1I_2,
$$

where

$$
(10.3)\qquad \begin{aligned}
A &= \int_N d\xi^1 d\xi^2, & I_1 &= \int_N (\xi^2)^2 d\xi^1 d\xi^2,\\
T &= \frac{1}{1+\sigma}\int_N \chi\, d\xi^1 d\xi^2, & I_2 &= \int_N (\xi^1)^2 d\xi^1 d\xi^2,
\end{aligned}
$$

N being a cross-section of the strained rod. We note that

$$
(10.4)\qquad A\epsilon^2 = A', \qquad I_\alpha\epsilon^4 = I'_\alpha, \qquad TE\epsilon^4 = T',
$$

where A' is the area of $\overline{N}$ (Figure 1), I'_α are the principal moments of inertia of $\overline{N}$, T' is the torsional rigidity of a straight rod with $\overline{N}$ as cross-section, E is Young's modulus and ϵ is defined in (7.1).

The first order macroscopic equations of equilibrium (8.14) and (8.15) are satisfied if

$$
(10.5)\qquad \frac{dB}{dx^0} = 0,
$$

$$
(10.6)\qquad (\overset{(\alpha)}{\bar{g}_{00}} + A_\alpha)B = 0.
$$

Thus we must consider in turn the three classes of problems which arise according as the arbitrary functions A_α and B of x^0 satisfy the following conditions:

(i) $B=0$, $(\overset{(\alpha)}{\bar{g}_{00}}+A_\alpha)(\overset{(\alpha)}{\bar{g}_{00}}+A_\alpha)\neq 0$,

(ii) $B=\mathit{constant}\neq 0$, $\overset{(\alpha)}{\bar{g}_{00}}+A_\alpha=0$,

(iii) $B=0$, $\overset{(\alpha)}{\bar{g}_{00}}+A_\alpha=0$.

By (9.17) and (9.19) we see that, in Class (i) the curvature of M (the strained line of centroids) is finite ($O(\epsilon^0)$) and the elongation per unit length of M is

relatively small ($O(\epsilon^2)$). In Class (ii), the curvature is small ($O(\epsilon)$) and the elongation per unit length is relatively large ($O(\epsilon)$). In Class (iii), the curvature is small ($O(\epsilon)$) and the elongation per unit length is relatively small ($O(\epsilon^2)$).

Classes (ii) and (iii) deal with the straightening of thin rods by terminal force-systems (see §13). By "straightening" we mean "reducing the curvature of the strained line of centroids from order ϵ^0 to order ϵ."

CLASS (i). $B=0$, $(\bar{g}_{00}^{(\alpha)}+A_\alpha)(\bar{g}_{00}^{(\alpha)}+A_\alpha)\neq 0$ (*elongation per unit length relatively small, curvature finite*).

The equations (10.1) and (10.2) reduce to

$$
\begin{aligned}
&P_{i(3)} = 0, && Q_{0(4)} = DT,\\
&Q_{1(4)} = \tfrac{1}{2}A_2 I_1, && Q_{2(4)} = -\tfrac{1}{2}A_1 I_2,
\end{aligned}
\tag{10.7}
$$

and the macroscopic equations of equilibrium of the second order (8.16) and (8.17) take the form

$$
\begin{aligned}
&\frac{dP_{0(4)}}{dx^0} + \tfrac{1}{2}(\bar{g}_{00}^{(\alpha)} + A_\alpha)P_{\alpha(4)} = 0,\\
&\frac{dP_{\alpha(4)}}{dx^0} - \tfrac{1}{2}(\bar{g}_{00}^{(\alpha)} + A_\alpha)P_{0(4)} - \tfrac{1}{2}(2D + \bar{g}_{02}^{(1)} - \bar{g}_{01}^{(2)})c_{\alpha\beta}P_{\beta(4)} = 0,\\
&\frac{dQ_{0(4)}}{dx^0} + \tfrac{1}{2}(\bar{g}_{00}^{(\alpha)} + A_\alpha)Q_{\alpha(4)} = 0,\\
&\frac{dQ_{\alpha(4)}}{dx^0} - \tfrac{1}{2}(\bar{g}_{00}^{(\alpha)} + A_\alpha)Q_{0(4)} - \tfrac{1}{2}(2D + \bar{g}_{02}^{(1)} - \bar{g}_{01}^{(2)})c_{\alpha\beta}Q_{\beta(4)} - c_{\alpha\beta}P_{\beta(4)} = 0.
\end{aligned}
\tag{10.8}
$$

Thus—

IV. In Class (i) the external terminal force-systems, which act only on the ends of the rod, are of order ϵ^4 or higher. The nine unknown functions A_α, D, $P_{i(4)}$, $Q_{i(4)}$, of x^0, can be determined from (10.7) and (10.8). Once A_α and D have been found in this way, the principal parts of the rotation $\omega^i_{(0)}$, the reduced stress tensor $T_{ij(1)}$, the metric tensor of the strained rod $g_{ij(1)}$ and the strain tensor $e_{ij(1)}$ can be found from (9.17), (9.15), (9.16) and (7.14).

From (7.36) and (9.17), we note that

$$
\begin{aligned}
D &= \overset{0}{\omega}_{(0)} - \bar{\overset{0}{\omega}},\\
A_\alpha &= -2c_{\alpha\beta}(\overset{\beta}{\omega}_{(0)} - \bar{\overset{\beta}{\omega}}),
\end{aligned}
\tag{10.9}
$$

and consequently (10.7) and (10.8), when expressed entirely in terms of the small quantities, become

$$
\begin{aligned}
&Q_0 = (\omega^0 - \bar{\omega}^0)T'/E,\\
&Q_1 = (\omega^1 - \bar{\omega}^1)I_1', \quad Q_2 = (\omega^2 - \bar{\omega}^2)I_2',
\end{aligned}
\tag{10.10}
$$

$$\frac{dP_i}{dx^0} + c^{ijk}\omega^j P_k = 0, \qquad \frac{dQ_i}{dx^0} + c^{ijk}\omega^j Q_k - c^{0ij}P_j = 0, \tag{10.11}$$

where EP_i and EQ_i are respectively the force and couple statically equivalent to the reaction across a cross-section of the strained rod, E, T' and I_α' are as defined in (10.4), $\bar{\omega}^i$ and ω^i are respectively the rotations in the unstrained and strained rods, c^{ijk} is a permutation symbol and x^0 is the arc length measured along the unstrained line of centroids. We note that (10.10) and (10.11) are a system of nine equations from which we are to determine the nine unknowns P_i, Q_i and ω^i.

In previous treatments of the finite displacement of thin rods the equations (10.10) and (10.11) have been deduced, but those treatments do not go past this point and are thus equivalent to the first stage of the systematic ϵ-method of approximation developed for problems of Class (i) only.

CLASS (ii). $B = constant \neq 0$, $\bar{g}_{00}^{(\alpha)} + A_\alpha = 0$ *(elongation per unit length relatively large, curvature small).*

The equations (10.1) and (10.2) reduce to

$$\begin{gathered} P_{0(3)} = \tfrac{1}{2}BA, \qquad P_{\alpha(3)} = 0, \\ Q_{0(4)} = DT, \qquad Q_{1(4)} = -\tfrac{1}{2}\overset{(2)}{\bar{g}}_{00} I_1, \qquad Q_{2(4)} = \tfrac{1}{2}\overset{(1)}{\bar{g}}_{00} I_2. \end{gathered} \tag{10.12}$$

The second order macroscopic equations of equilibrium (8.16) and (8.17) are found to involve $(g_{00(2),\alpha})_C$ which, by the second order equations of compatibility (8.7), have the forms

$$(g_{00(2),\alpha})_C = 2M_\alpha, \tag{10.13}$$

M_α being an arbitrary function of x^0. By integration of these second order macroscopic equations of equilibrium, and by (10.12), it is then found that

$$\begin{aligned} P_{0(3)} &= \text{constant}, \qquad P_{0(4)} = \text{constant}, \\ P_{1(4)} &= -\tfrac{1}{2}\overset{(1)}{\bar{g}}_{00,0} I_2 + \tfrac{1}{4}(2D + \overset{(1)}{\bar{g}}_{02} - \overset{(2)}{\bar{g}}_{01})\overset{(2)}{\bar{g}}_{00} I_1 + (\chi_{,2})_C P_{0(3)} D, \\ P_{2(4)} &= -\tfrac{1}{2}\overset{(2)}{\bar{g}}_{00,0} I_1 + \tfrac{1}{4}(-2D + \overset{(2)}{\bar{g}}_{01} - \overset{(1)}{\bar{g}}_{02})\overset{(1)}{\bar{g}}_{00} I_2 - (\chi_{,1})_C P_{0(3)} D, \\ Q_{0(4)} &= DT = \text{constant}, \\ Q_{1(4)} &= -\tfrac{1}{2}\overset{(2)}{\bar{g}}_{00} I_1, \qquad Q_{2(4)} = \tfrac{1}{2}\overset{(1)}{\bar{g}}_{00} I_2, \end{aligned} \tag{10.14}$$

$$\begin{aligned} M_1 P_{0(3)} = &-\tfrac{1}{2}\overset{(1)}{\bar{g}}_{00,00} I_2 + \tfrac{1}{2}(2D + \overset{(1)}{\bar{g}}_{02} - \overset{(2)}{\bar{g}}_{01})\overset{(2)}{\bar{g}}_{00,0} I_1 \\ &+ \tfrac{1}{8}(2D + \overset{(1)}{\bar{g}}_{02} - \overset{(2)}{\bar{g}}_{01})\overset{(1)}{\bar{g}}_{00} I_2 + \tfrac{1}{4}(\overset{(1)}{\bar{g}}_{02,0} - \overset{(2)}{\bar{g}}_{01,0})\overset{(2)}{\bar{g}}_{00} I_1, \\ M_2 P_{0(3)} = &-\tfrac{1}{2}\overset{(2)}{\bar{g}}_{00,00} I_1 + \tfrac{1}{2}(-2D + \overset{(2)}{\bar{g}}_{01} - \overset{(1)}{\bar{g}}_{02})\overset{(1)}{\bar{g}}_{00,0} I_2 \\ &+ \tfrac{1}{8}(-2D + \overset{(2)}{\bar{g}}_{01} - \overset{(1)}{\bar{g}}_{02})\overset{(2)}{\bar{g}}_{00} I_1 + \tfrac{1}{4}(\overset{(2)}{\bar{g}}_{01,0} - \overset{(1)}{\bar{g}}_{02,0})\overset{(1)}{\bar{g}}_{00} I_2. \end{aligned} \tag{10.15}$$

The arbitrary functions M_α of x^0 in (10.13) are given by (10.15). We note that D must be constant. From (9.17) and (9.18) we obtain

$$\overset{0}{\omega}_{(0)} = \tfrac{1}{2}c_{\alpha\beta}\overset{(\alpha)}{\bar g}_{0\beta} + D, \qquad \overset{\alpha}{\omega}_{(1)} = c_{\alpha\beta}M_\alpha, \tag{10.16}$$

giving the principal parts of the twist and curvature. Thus—

V. In Class (ii), there act on the ends of the rod arbitrary equal and opposite forces of order ϵ^3, arbitrary twisting couples of order ϵ^4 and particular bending couples of order ϵ^4 given by (10.14) and depending only on the configuration of the unstrained rod. The principal parts of the reaction across a cross-section satisfy (10.14). The principal parts of the twist and curvature of the strained line of centroids, of order ϵ^0 and ϵ respectively, are given by (10.16). The principal parts of the reduced stress tensor T_{ij}, the metric tensor g_{ij} and the strain tensor e_{ij} can be found from (9.15), (9.16) and (7.14), respectively.

Class (iii). $B=0$, $\bar g_{00}^{(\alpha)}+A_\alpha=0$ (*elongation per unit length relatively small, curvature small*).

The equations of equilibrium (10.1) and (10.2) reduce to

$$P_{i(3)} = 0, \tag{10.17}$$

$$Q_{0(4)} = DT, \qquad Q_{1(4)} = -\tfrac{1}{2}\overset{(2)}{\bar g}_{00} I_1, \qquad Q_{2(4)} = \tfrac{1}{2}\overset{(1)}{\bar g}_{00} I_2. \tag{10.18}$$

The three equations in (10.18) and the six second order macroscopic equations of equilibrium (8.16) and (8.17) are to be solved for the seven unknowns $Q_{i(4)}$, $P_{i(4)}$, D. Solutions exist only if

$$D = 0, \tag{10.19}$$

$$\begin{aligned} &-\overset{(1)}{\bar g}_{00,00}I_2 + \tfrac{1}{2}(\overset{(1)}{\bar g}_{02,0} - \overset{(2)}{\bar g}_{01,0})\overset{(2)}{\bar g}_{00} I_1 \\ &\qquad + (\overset{(1)}{\bar g}_{02} - \overset{(2)}{\bar g}_{01})\overset{(2)}{\bar g}_{00,0}I_1 + \tfrac{1}{4}(\overset{(1)}{\bar g}_{02} - \overset{(2)}{\bar g}_{01})^2\overset{(1)}{\bar g}_{00} I_2 = 0, \\ &-\overset{(2)}{\bar g}_{00,00}I_1 + \tfrac{1}{2}(\overset{(2)}{\bar g}_{01,0} - \overset{(1)}{\bar g}_{02,0})\overset{(1)}{\bar g}_{00} I_2 \\ &\qquad + (\overset{(2)}{\bar g}_{01} - \overset{(1)}{\bar g}_{02})\overset{(1)}{\bar g}_{00,0}I_2 + \tfrac{1}{4}(\overset{(2)}{\bar g}_{01} - \overset{(1)}{\bar g}_{02})^2\overset{(2)}{\bar g}_{00} I_1 = 0; \end{aligned} \tag{10.20}$$

(10.20) can also be written in the form

$$\begin{aligned} &(\overset{0}{\bar\omega}_{,0}\overset{1}{\bar\omega} + 2\overset{0}{\bar\omega}\overset{1}{\bar\omega}_{,0})I_1/I_2 + \overset{2}{\bar\omega}_{,00} - (\overset{0}{\bar\omega})^2\overset{2}{\bar\omega} = 0, \\ &(\overset{0}{\bar\omega}_{,0}\overset{2}{\bar\omega} + 2\overset{0}{\bar\omega}\overset{2}{\bar\omega}_{,0})I_2/I_1 - \overset{1}{\bar\omega}_{,00} + (\overset{0}{\bar\omega})^2\overset{1}{\bar\omega} = 0, \end{aligned} \tag{10.21}$$

in which I_1/I_2 is the ratio of the principal moments of inertia of the cross-section, $\bar\omega^i$ is the rotation vector defined in §2, and the subscript 0 preceded by a comma denotes differentiation with respect to x^0, the arc length of the line of centroids. When (10.21) are satisfied, the solutions of (10.17), (10.18), (10.19), (8.16) and (8.17) are

$$(10.22)\qquad \begin{aligned} P_{0(4)} &= \text{constant}, \\ P_{1(4)} &= -\tfrac{1}{2}\overset{(1)}{\bar{g}}_{00,0}I_2 + \tfrac{1}{4}(\overset{(1)}{\bar{g}}_{02} - \overset{(2)}{\bar{g}}_{01})\overset{(2)}{\bar{g}}_{00}I_1, \\ P_{2(4)} &= -\tfrac{1}{2}\overset{(2)}{\bar{g}}_{00,0}I_1 + \tfrac{1}{4}(\overset{(2)}{\bar{g}}_{01} - \overset{(1)}{\bar{g}}_{02})\overset{(1)}{\bar{g}}_{00}I_2, \\ Q_{0(4)} &= 0, \\ Q_{1(4)} &= -\tfrac{1}{2}\overset{(2)}{\bar{g}}_{00}I_1, \\ Q_{2(4)} &= \tfrac{1}{2}\overset{(1)}{\bar{g}}_{00}I_2, \end{aligned}$$

and we then have

$$(10.23)\qquad \overset{0}{\omega}_{(1)} = \overset{0}{\bar{\omega}}, \qquad \overset{\alpha}{\omega}_{(1)} = c_{\alpha\beta}M_{\alpha},$$

for the principal parts of the twist and curvature. Thus—

VI. Class (iii) deals only with those rods which satisfy (10.20). There act on the ends of the rod arbitrary forces of order ϵ^4, arbitrary twisting couples of order ϵ^5 or higher, and particular bending couples of order ϵ^4 given by (10.22) and depending only on the configuration of the unstrained rod. The principal parts of the reaction across a cross-section are given by (10.22). From (10.23) we see that the principal part of the twist ω^0 (of order ϵ^0) is equal to the twist $\bar{\omega}^0$ of the unstrained rod, and the principal parts of the curvature ω^α (of order ϵ) depend on M_α which are arbitrary functions of x^0 and cannot be determined from the equations of the first approximation. The principal parts of the reduced stress tensor T_{ij}, the metric tensor g_{ij} and the strain tensor e_{ij} are given by (9.15), (9.16) and (7.14) with $B=D=0$,

$$A_\alpha = -\bar{g}_{00}^{(\alpha)}.$$

If $\bar{\omega}^0=0$ (i.e., there is no twist in the unstrained rod), (10.20) reduce to

$$(10.24)\qquad \overset{(\alpha)}{\bar{g}}_{00,00} = 0.$$

By integrating, we obtain

$$(10.25)\qquad \overset{(\alpha)}{\bar{g}}_{00} = V_\alpha \overset{0}{x} + W_\alpha,$$

V_α, W_α being arbitrary constants. Thus by (7.36),

$$(10.26)\qquad \bar{\omega}^\alpha = \tfrac{1}{2}c_{\alpha\beta}(V_\beta x^0 + W_\beta),$$

i.e., the curvature of the unstrained line of centroids is directly proportional to its arc length.

11. **The second approximation: the stress.** As indicated in §8, II, the second order equations of equilibrium (8.3) and the second order equations of compatibility (8.7), (8.8), (8.9) with the second order boundary conditions (8.13) are to be solved for $T_{ij(2)}$ and $g_{ij(2)}$, which are themselves related by the second order stress-strain relations (8.11). These equations are very complicated. We shall now determine $T_{ij(2)}$ and $g_{ij(2)}$ when the cross-section of the unstrained rod is a circle of radius d, the unstrained line of centroids is a

plane curve with curvature proportional to its arc length, and the force-systems acting on the ends of the rod are such that the line of centroids remains in its original plane. The problem is thus two-dimensional and belongs to Class (iii). We choose the axis of x^1 perpendicular to the plane of the line of centroids. Thus from (10.3) we find that

$$A = \pi R^2, \qquad I_1 = I_2 = \tfrac{1}{4}\pi R^4, \tag{11.1}$$

where $\epsilon R = d$. Also

$$\bar{\omega}^0 = 0, \qquad \bar{\omega}^1 = \tfrac{1}{2}(V_2 x^0 + W_2), \qquad \bar{\omega}^2 = 0, \tag{11.2}$$

$$\bar{g}_{\alpha\beta} = \overset{\alpha}{\delta_\beta}, \qquad \bar{g}_{0\alpha} = 0, \qquad \bar{g}_{00} = [1 + \tfrac{1}{2}\overset{2}{x}(V_2 \overset{0}{x} + W_2)]^2, \tag{11.3}$$

$$P_1 = Q_0 = Q_2 = 0, \tag{11.4}$$

$$P_{0(4)} = p, \qquad P_{2(4)} = -\tfrac{1}{8}\pi R^4 V_2, \qquad Q_{1(4)} = -\tfrac{1}{8}\pi R^4(V_2 x^0 + W_2), \tag{11.5}$$

$$T_{\alpha\beta(1)} = 0, \qquad T_{0\alpha(1)} = 0, \qquad T_{00(1)} = -\tfrac{1}{2}\xi^2(V_2 x^0 + W_2), \tag{11.6}$$

$$g_{\alpha\beta(1)} = \sigma\overset{\alpha}{\delta_\beta}\overset{2}{\xi}(V_2 \overset{0}{x} + W_2), \qquad g_{0\alpha(1)} = 0, \qquad g_{00(1)} = 0, \tag{11.7}$$

p being an arbitrary constant and V_2 and W_2 known constants. We find that the second order equations of equilibrium (8.3) reduce to

$$T_{0\alpha(2),\alpha} - \tfrac{1}{2}V_2\xi^2 = 0, \qquad T_{\alpha\beta(2),\beta} = 0, \tag{11.8}$$

the second order equations of compatibility (8.7), (8.8) and (8.9) to

$$\begin{aligned} &g_{00(2),\alpha\beta} = 0, \\ &g_{01(2),22} - g_{02(2),12} = 0, \qquad g_{02(2),11} - g_{01(2),21} + \sigma V_2 = 0, \\ &2g_{12(2),12} - g_{11(2),22} - g_{22(2),11} + \sigma^2(V_2 x^0 + W_2)^2 = 0, \end{aligned} \tag{11.9}$$

the second order boundary conditions (8.13) to

$$T_{i\beta(2)} n_{\beta(0)} = 0, \tag{11.10}$$

and the second order stress-strain relations (8.11) to

$$\begin{aligned} g_{\alpha\beta(2)} &= -2\sigma\overset{\alpha}{\delta_\beta} T_{ii(2)} + 2(1+\sigma)T_{\alpha\beta(2)} \\ &\quad + (\overset{2}{\sigma} - \tfrac{1}{2}\sigma_1)\overset{\alpha}{\delta_\beta}(\overset{2}{\xi})^2(V_2 \overset{0}{x} + W_2)^2, \\ g_{0\alpha(2)} &= 2(1+\sigma)T_{0\alpha(2)}, \\ g_{00(2)} &= -2\sigma T_{ii(2)} + 2(1+\sigma)T_{00(2)} + \tfrac{1}{4}\Lambda(\xi^2)^2(V_2 x^0 + W_2)^2, \end{aligned} \tag{11.11}$$

where

$$\Lambda = 7 + 10\sigma + 4\sigma^2 - 2\sigma_1 + 2\sigma_2, \tag{11.12}$$

and σ_1, σ_2 and σ are the elastic constants introduced in §4. By solving these equations we find that

$$T_{00(2)} = \frac{1}{32}\left[\frac{\Gamma\sigma}{1-\sigma}(R^2 - 2\xi^\alpha\xi^\alpha) - 4\Lambda\xi^2\xi^2\right](V_2x^0 + W_2)^2 + M_2\xi^2 + \tfrac{1}{2}L,$$

$$(11.13)\quad T_{\alpha\beta(2)} = \frac{\Gamma}{64(1-\sigma)}[(R^2 - \xi^\gamma\xi^\gamma)\delta^\alpha_\beta - 2c_{\alpha\gamma}c_{\beta\delta}\xi^\gamma\xi^\delta](V_2x^0 + W_2)^2,$$

$$T_{0\alpha(2)} = \frac{1+2\sigma}{8(1+\sigma)}\xi^1 c_{\alpha\beta}\xi^\beta V_2 - \frac{3+2\sigma}{16(1+\sigma)}(R^2 - \xi^\gamma\xi^\gamma)\delta^2_\alpha V_2,$$

$$g_{00(2)} = 2M_2\xi^2 + L,$$

$$g_{\alpha\beta(2)} = \left\{\frac{(1+\sigma)}{32(1-\sigma)}[(1-2\sigma)R^2 - (1-4\sigma)\xi^\gamma\xi^\gamma]\right.$$

$$(11.14)\qquad \left. + \tfrac{1}{4}(\Lambda\sigma + 4\sigma^2 - 2\sigma_1)\xi^2\xi^2\right\}\delta^\alpha_\beta(V_2x^0 + W_2)^2 - \sigma\delta^\alpha_\beta(2M_2\xi^2 + L)$$

$$- \frac{\Gamma(1+\sigma)}{16(1-\sigma)}c_{\alpha\gamma}c_{\beta\delta}\xi^\gamma\xi^\delta(V_2x^0 + W_2)^2,$$

$$g_{0\alpha(2)} = \tfrac{1}{4}(1+2\sigma)\xi^1 c_{\alpha\beta}\xi^\beta V_2 - \tfrac{1}{8}(3+2\sigma)(R^2 - \xi^\gamma\xi^\gamma)\delta^2_\alpha V_2;$$

$c_{\alpha\beta}$ is a permutation symbol; M_2 and L are arbitrary functions of x^0; V_2 and W_2 are the constants occurring in (11.2); $R=d/\epsilon$, d being the radius of the cross-section of the unstrained rod; σ, σ_1, σ_2 and σ_3 are elastic constants, Λ is given by (11.12) and

$$(11.15)\qquad \Gamma = (\Lambda\sigma + 2\sigma^2 - 2\sigma_1)/(1+\sigma).$$

By (10.23), the principal part of the curvature of the strained line of centroids is given by

$$(11.16)\qquad \overset{1}{\omega}_{(1)} = M_2.$$

Also, by (7.42) the principal part of the elongation per unit length of the line of centroids is given by

$$(11.17)\qquad e_{(2)} = \tfrac{1}{2}L.$$

12. The second approximation: the macroscopic equations of equilibrium. The next step is the determination of the arbitrary functions M_2 and L of x^0 occurring in (11.13) and (11.14) by means of the third order macroscopic equations of equilibrium (8.18) and (8.19). Now (7.33) and (7.35) reduce to

$$P_{0(4)} = \tfrac{1}{2}\pi R^2 L - \tfrac{1}{32}\pi R^4(\Lambda + 4\sigma)(V_2x^0 + W_2)^2,$$
$$(12.1)\qquad P_{2(4)} = -\tfrac{1}{8}\pi R^4 V_2,$$
$$Q_{1(5)} = \tfrac{1}{4}\pi R^4 M_2,$$

and by comparing (12.1) with (11.5), we see that

$$(12.2)\qquad L = \tfrac{1}{16}\pi R^2(\Lambda + 4\sigma)(V_2x^0 + W_2)^2 + 2p\pi^{-1}R^{-2},$$

p being an arbitrary constant. The third order macroscopic equations of equilibrium (8.18) and (8.19) reduce to

$$(12.3)\qquad P_{0(5),0} + M_2P_{2(4)} = 0,\quad P_{2(5),0} - M_2P_{0(4)} = 0,\quad Q_{1(5),0} - P_{2(5)} = 0,$$

and by solving (12.1) and (12.3), we find that

$$(12.4)\qquad \begin{aligned} &P_{0(4)} = p,\quad P_{2(4)} = -\tfrac{1}{8}\pi R^4V_2,\quad Q_{1(4)} = -\tfrac{1}{8}\pi R^4(V_2x^0 + W_2),\\ &P_{0(5)} = V_2u \cosh nx^0 + V_2vn^{-1} \sinh nx^0 + t,\\ &P_{2(5)} = 2n^2u \cosh nx^0 + 2nv \sinh nx^0,\\ &Q_{1(5)} = 2nu \sinh nx^0 + 2v \cosh nx^0, \end{aligned}$$

$$(12.5)\qquad M_2 = 8\pi^{-1}R^{-4}(un \sinh nx^0 + v \cosh nx^0),$$

where p, t, u and v are arbitrary constants and

$$(12.6)\qquad n = 2R^{-2}(p/\pi)^{1/2}.$$

We assign $P_{0(4)}$, $P_{0(5)}$, $P_{2(5)}$ and $Q_{1(5)}$ arbitrarily at one end of the rod and hence determine p, t, u and v.

From (11.16) and (11.17) we find that the principal parts of the curvature and the elongation per unit length are given by

$$(12.7)\qquad \overset{1}{\omega}_{(1)} = 8\overset{-1}{\pi}\overset{-4}{R}(un \sinh n\overset{0}{x} + v \cosh n\overset{0}{x}),$$

$$(12.8)\qquad e_{(2)} = \tfrac{1}{32}R^4(\Lambda + 4\sigma)(V_2x^0 + W^2)^2 + p\pi^{-1}R^{-2}.$$

13. **The straightening of naturally curved thin rods by terminal force-systems.** A thin rod is said to be "straightened" when the curvature of the line of centroids is reduced from order ϵ^0 (finite) to order ϵ (small). In §10 it is seen how all problems in which thin rods undergo finite displacement with small strain can be divided into three classes. Classes (ii) and (iii) deal exclusively with "straightening," and will be considered in turn.

Straightening as a problem of Class (ii). The theory pertaining to problems of Class (ii) is developed in §10, from which we obtain the result:

VII. Any thin rod with uniform cross-section and finite curvature can be straightened without undergoing finite strain by terminal force-systems consisting of arbitrary equal and opposite forces of order ϵ^3 and particular bending couples of order ϵ^4. The first approximation yields the residue curvature of the strained line of centroids. If the curvature of the unstrained line of centroids vanishes at the ends, no terminal bending couples are necessary. Further, if the twist and rate of change of curvature both vanish at the ends of the rod, the lines of action of the two forces acting one on each end will coincide and be tangent to the strained line of centroids at its ends.

As an example of a problem of Class (ii), we shall now straighten the rod

shown in Figure 4. The unstrained line of centroids $\overline{M}$ is that part of an ellipse lying between the ends of the major axis, a and b are the semi-axes of this ellipse and rectangular cartesian coordinates x and y are taken as shown. The cross-section is assumed to be symmetrical with respect to the plane of $\overline{M}$,

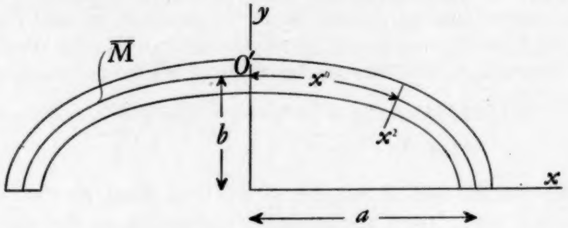


FIG. 4. An unstrained rod; the line of centroids $\overline{M}$ is the arc of an ellipse lying between the ends of the major axis

the axis of x^1 is taken perpendicular to this plane and x^0 is the arc length of $\overline{M}$ measured from its middle point O'. If the equation of $\overline{M}$ is written in the form

$$x = a \operatorname{sn} \zeta, \qquad y = b \operatorname{cn} \zeta, \tag{13.1}$$

then

$$x^0 = a \int_0^\zeta \operatorname{dn}^2 \zeta \, d\zeta, \tag{13.2}$$

ζ being a parameter which vanishes at O', and $\operatorname{sn} \zeta$, $\operatorname{cn} \zeta$ and $\operatorname{dn} \zeta$ the Jacobian elliptic functions of modulus k $(=(1-b^2/a^2)^{1/2})$. We then have

$$\bar{\omega}^0 = 0, \qquad \bar{\omega}^1 = a^{-2}b/\operatorname{dn}^3 \zeta, \qquad \bar{\omega}^2 = 0, \tag{13.3}$$

$\bar{\omega}^i$ being the rotation vector introduced in §2 ($\bar{\omega}^1$ is now the curvature of $\overline{M}$). From (7.36) we then find that

$$\overset{(1)}{\bar{g}_{02}} = \overset{(2)}{\bar{g}_{01}}, \qquad \overset{(1)}{\bar{g}_{00}} = 0, \qquad \overset{(2)}{\bar{g}_{00}} = -\overset{-2}{a} b/\overset{3}{\operatorname{dn}} \zeta, \tag{13.4}$$

whence (10.14), (10.15) and (10.16) reduce to

$$\begin{aligned} &P_{0(3)} = \text{constant}, \qquad P_{2(4)} = 3I_1 a^{-3} b k^2 \operatorname{sn} \zeta \operatorname{cn} \zeta/\operatorname{dn}^6 \zeta, \\ &Q_{1(4)} = I_1 a^{-2} b/\operatorname{dn}^3 \zeta, \end{aligned} \tag{13.5}$$

$$\overset{1}{\omega}_{(1)} = \frac{3I_1k^2}{P_{0(3)}a^8}\left(\frac{\operatorname{cn}^2 \zeta - \operatorname{sn}^2 \zeta}{\operatorname{dn}^7 \zeta} + 6k^2 \cdot \frac{\operatorname{sn}^2 \zeta \operatorname{cn}^2 \zeta}{\operatorname{dn}^9 \zeta}\right). \tag{13.6}$$

Thus the principal part of the reaction across a general cross-section of the strained rod consists of a tension P_0' $(=EP_{0(3)}\epsilon^3)$, a shearing force P_2' $(=EP_{2(4)}\epsilon^4)$ and a bending couple Q_1' $(=EQ_{1(4)}\epsilon^4)$, whence

$$
(13.7)\qquad
\begin{aligned}
P_0' &= \text{an arbitrary constant of order } \epsilon^3,\\
P_2' &= 3EI_1' a^{-3}bk^2 \operatorname{sn}\zeta \operatorname{cn}\zeta/\operatorname{dn}^6\zeta,\\
Q_1' &= EI_1' a^{-2}b/\operatorname{dn}^3\zeta,
\end{aligned}
$$

I_1' being the moment of inertia of the unstrained cross-section about an axis through its centroid and perpendicular to the plane of $\overline{M}$, and ζ a parameter related by (13.2) to the arc length x^0 of the unstrained rod. Also, the principal part of the curvature of the strained line of centroids is $\epsilon\omega_{(1)}^1$ or

$$
(13.8)\qquad \frac{3EI_1' bk^2}{P_0' a^8}\left(\frac{\operatorname{cn}^2\zeta - \operatorname{sn}^2\zeta}{\operatorname{dn}^7\zeta} + 6k^2\,\frac{\operatorname{sn}^2\zeta \operatorname{cn}^2\zeta}{\operatorname{dn}^9\zeta}\right).
$$

At the right-hand end of the rod, $\operatorname{sn}\zeta=1$, $\operatorname{cn}\zeta=0$, $\operatorname{dn}\zeta=b/a$, and thus $P_2'=0$, $Q_1'=EI_1'ab^{-2}$. Thus the force-system acting on the right-hand end of the rod consists of an arbitrary force P_0' of order ϵ^3 and a particular bending couple, as shown in Figure 5. From symmetry, a similar force-system acts on the other end of the rod. Further, since $P_2'=0$ at the ends of the rod, the lines of action of the two forces P_0' coincide and are tangent to the strained line of centroids at its ends.

FIG. 5. The rod in Figure 4, straightened

At O', $\operatorname{dn}\zeta=1$ and hence $Q_1'=EI_1'a^{-2}b$. Thus by taking the moment about O' of the forces and couples acting on the right-hand half of the rod, we find that the distance d' (Figure 5) is

$$
(13.9)\qquad d' = \frac{EI_1'(a^3-b^3)}{P'a^2b^2}\,\cdot
$$

Straightening as a problem of Class (iii). As mentioned previously, all problems of Class (iii) deal with the straightening of thin rods by terminal force-systems. From the theory developed in §§10–12 for problems of this class, it can be conclude that—

VIII. Any thin rod of uniform cross-section and finite curvature, and satisfying (10.21), can be straightened without undergoing finite strain by means of terminal force-systems consisting of an arbitrary force and a particular bending couple, both of order ϵ^4. The residue curvature of the strained line of centroids can be found only by recourse to the second approximation.

As an example of a problem of Class (iii), we shall now straighten the rod shown in Figure 6. The cross-section is a circle of radius d. The unstrained line

of centroids $\overline{M}$ is a semicircle of radius c and x^0 is its arc length measured from its middle point O'. The x^1-axis is taken perpendicular to the plane of $\overline{M}$. Thus

$$(13.10) \qquad \bar{\omega}^0 = 0, \qquad \bar{\omega}^1 = -1/c, \qquad \bar{\omega}^2 = 0,$$

$\bar{\omega}^i$ being the rotation vector introduced in §2. Hence by comparison with (11.2),

$$(13.11) \qquad V_1 = V_2 = W_1 = 0, \qquad W_2 = -2/c,$$

and since the strained line of centroids M is to be a plane curve symmetrical

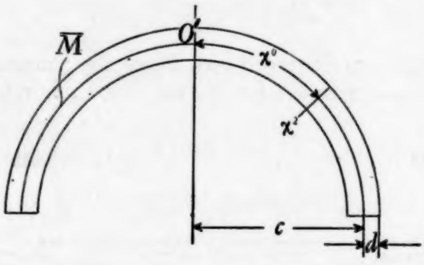


FIG. 6. An unstrained rod; the cross-section is circular and the line of centroids $\overline{M}$ is a semicircle

about its middle point O' (Figure 7), we find that (12.4), (12.7) and (12.8) reduce to

$$(13.12) \qquad \begin{aligned} &P_{0(4)} = p, \qquad Q_{1(4)} = \tfrac{1}{4}\pi R^4/c, \\ &P_{0(5)} = t, \qquad P_{2(5)} = 2vn \sinh nx^0, \qquad Q_{1(5)} = 2v \cosh nx^0, \end{aligned}$$

$$(13.13) \qquad \overset{1}{\omega}_{(1)} = \frac{8v}{\pi R^4} \cosh nx^0,$$

$$(13.14) \qquad e_{(2)} = (\Lambda + 4\sigma)\frac{R^2}{8c^2} + \frac{p}{\pi R^2},$$

where p is an arbitrary constant and $n = 2\,(p/\pi R^4)^{1/2}$. Thus the reaction across a cross-section of the strained rod consists of a normal force $P_0' + P_0''$ ($P_0' = EP_{0(4)}\epsilon^4 = Ep\epsilon^4$, $P_0'' = EP_{0(5)}\epsilon^5$), a shearing force P_2'' $(= EP_{2(5)}\epsilon^5)$, and a bending couple $Q_1' + Q_1''$ ($Q_1' = EQ_{1(4)}\epsilon^4$, $Q_1'' = EQ_{1(5)}\epsilon^5$); and

$$(13.15) \qquad \begin{aligned} P_0' &= \text{an arbitrary constant of order } \epsilon^4, \\ P_0'' &= \text{an arbitrary constant of order } \epsilon^5, \\ P_2'' &= 2Ev'' \sinh nx^0, \\ Q_1' &= \tfrac{1}{4}\pi E d^4/c, \qquad Q_1'' = 2Ev'' \cosh nx^0, \end{aligned}$$

where $n=2(P_0'/\pi Ed^4)^{1/2}$ and v'' is an arbitrary constant $(=\epsilon^5 v)$. In (13.15), the quantities with a single prime are of order ϵ^4, while those with a double prime are of order ϵ^5. The strained line of centroids M has a curvature ω^1 $(=\epsilon\omega^1_{(1)})$ given by

$$\omega^1 = \frac{8v''}{\pi d^4} \cosh nx^0, \tag{13.16}$$

and an elongation per unit length e $(=\epsilon^2 e_{(2)})$ given by

$$e = (\Lambda + 4\sigma)\frac{d^2}{8c^2} + \frac{P_0'}{\pi Ed^4}\cdot \tag{13.17}$$

Figure 7 shows the straightened rod. From the consideration just above of the reaction across a cross-section, we see that there must act on the ends

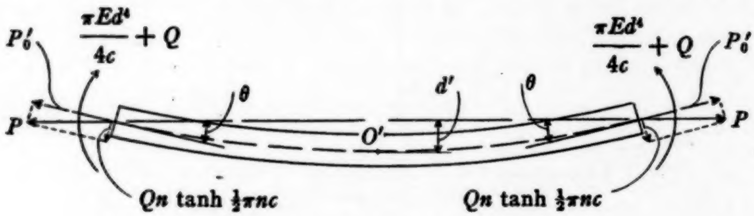


FIG. 7. The rod in Figure 6, straightened

of the rod equal and opposite forces P of order ϵ^4 and bending couples

$$\tfrac{1}{4}\pi Ed^4/c + Q$$

as shown, $\frac{1}{4}\pi Ed^4/c$ being of order ϵ^4 and Q of order ϵ^5. Further, since $x^0 = \pm\frac{1}{2}\pi c$ at the ends of the rod,

$$v'' = Q/(2E \cosh \tfrac{1}{2}\pi nc), \tag{13.18}$$

$$P\cos\theta = P_0', \qquad P\sin\theta = Qn\tanh\tfrac{1}{2}\pi nc, \tag{13.19}$$

θ being the small angle shown in Figure 7. Since θ is small, we set

$$\sin\theta = \theta, \qquad \cos\theta = 1.$$

Thus $P=P_0'$ approximately, and

$$n = 2(P/\pi Ed^4)^{1/2}, \tag{13.20}$$

$$\theta = \frac{Q}{P} n \tanh \tfrac{1}{2}\pi nc, \tag{13.21}$$

$$e = (\Lambda + 4\sigma)\frac{d^2}{8c^2} + \frac{P}{\pi Ed^4}\cdot \tag{13.22}$$

From (13.13) the curvature of the strained line of centroids is then

$$\omega^1 = \frac{4Q}{\pi d^4} \frac{\cosh nx^0}{\cosh \frac{1}{2}\pi nc}. \tag{13.23}$$

At O', $x^0 = 0$ and hence

$$Q_1'' = Q \operatorname{sech} \tfrac{1}{2}\pi nc.$$

Thus by taking the moment about O' of the forces and couples acting on the right-hand half of the rod, we find that the distance d' (Figure 7) is

$$d' = \frac{Q}{P}(1 - \operatorname{sech} \tfrac{1}{2}\pi nc). \tag{13.24}$$

If P vanishes (i.e., no external forces are applied to the ends of the rod), then $n = 0$ and

$$\theta = \lim_{P\to 0}\left(\frac{Q}{P} n \tanh \tfrac{1}{2}\pi nc\right) = 2Qc/(Ed^4), \tag{13.25}$$

$$e = (\Lambda + 4\sigma)d^2/(8c^2), \tag{13.26}$$

$$\omega^1 = 4Q/(\pi E d^4). \tag{13.27}$$

In this case the line of centroids of the strained rod is an arc of a circle of radius

$$\tfrac{1}{4}\pi E d^4/Q.$$

14. **Remarks on the experimental determination of elastic constants.** The stress-strain relations (4.3) used in this paper contain five independent elastic constants $E, \sigma, \sigma_1, \sigma_2, \sigma_3$; E is Young's modulus and σ is Poisson's ratio. Both E and σ can be determined for any isotropic elastic material by simple experiments involving the stretching and twisting of a straight rod. In this section, we shall see how a relation between σ_1 and σ_2 can be determined experimentally by straightening the thin rod shown in Figure 6. The straightened rod is shown in Figure 7. When terminal bending couples alone of magnitude $\frac{1}{4}\pi E d^4/c$ act on the rod, the elongation per unit length of the line of centroids is given by (13.26). By actually measuring this elongation, we can thus obtain an experimental value for $\Lambda + 4\sigma$ for the particular material of which the rod is constructed. Now σ is known for most elastic materials. Thus Λ can also be found. By (11.12), this gives a relation between σ_1 and σ_2. In order to determine the arithmetical values of the three elastic constants σ_1, σ_2 and σ_3 occurring in the stress-strain relations (4.3), it would be necessary to devise additional experimental methods of obtaining two additional relations between σ_1, σ_2 and σ_3.

The author would like to take this opportunity to thank Dr. J. L. Synge both for suggesting the problem considered in this paper and for much assistance in the development of the mathematical theory.

Appendix A. Proof of Saint-Venant's hypothesis for a thin rod displaced finitely[12]. We are to show that, given

$$T_{\alpha\beta(1),\beta} = 0, \tag{A.1}$$

$$g_{00(1),\alpha\beta} = 0, \tag{A.2}$$

$$2g_{12(1),12} - g_{11(1),22} - g_{22(1),11} = 0, \tag{A.3}$$

$$g_{\alpha\beta(1)} = \bar{g}_{\alpha\beta}^{(\gamma)}\xi^{\gamma} - 2\sigma\delta_{\beta}^{\alpha}T_{kk(1)} + 2(1+\sigma)T_{\alpha\beta(1)}, \tag{A.4}$$

with the boundary conditions

$$T_{\alpha\beta(1)}n_{\beta(0)} = 0, \tag{A.5}$$

it is necessary that $T_{\alpha\beta(1)} = 0$.

From (A.2),

$$g_{00(1)} = A'_{\alpha}\xi^{\alpha} + B, \tag{A.6}$$

A'_{α} and B being arbitrary functions of x^0. Then by (A.4),

$$T_{00(1)} = \sigma T_{\alpha\alpha(1)} + \tfrac{1}{2}(A_{\alpha}\xi^{\alpha} + B), \tag{A.7}$$

$$T_{ii(1)} = (1+\sigma)T_{\alpha\alpha(1)} + \tfrac{1}{2}(A_{\alpha}\xi^{\alpha} + B), \tag{A.8}$$

A_{α} being an arbitrary function of x^0 such that $A_{\alpha} = -\bar{g}_{00}^{(\alpha)} + A'_{\alpha}$. Substitution in (A.3) for $g_{\alpha\beta(1)}$ from (A.4) and then for $T_{ii(1)}$ from (A.8) gives

$$T_{11(1),22} + T_{22(1),11} - 2T_{12(1),12} = \sigma T_{\alpha\alpha(1),\beta\beta}. \tag{A.9}$$

It is necessary to show that there exist functions $u_{\alpha}(x^0, \xi^{\beta})$ and a constant σ' such that

$$\tfrac{1}{2}(u_{\alpha,\beta} + u_{\beta,\alpha}) = (1+\sigma')T_{\alpha\beta(1)} - \sigma'\delta_{\beta}^{\alpha}T_{\gamma\gamma(1)}. \tag{A.10}$$

These are satisfied when $\alpha=\beta=1$, $\alpha=\beta=2$ if

$$\begin{aligned} u_1 &= \int_0^{\xi^1} [(1+\sigma')T_{11(1)} - \sigma' T_{\gamma\gamma(1)}]d\xi^1 + u'_1(x^0, \xi^2), \\ u_2 &= \int_0^{\xi^2} [(1+\sigma')T_{22(1)} - \sigma' T_{\gamma\gamma(1)}]d\xi^2 + u'_2(x^0, \xi^1), \end{aligned} \tag{A.11}$$

and the third equation in (A.10) is satisfied if

$$\begin{aligned} u'_{1,2} + u'_{2,1} &+ \int_0^{\xi^1} [(1+\sigma')T_{11(1),2} - \sigma' T_{\gamma\gamma(1),2}]d\xi^1 \\ &+ \int_0^{\xi^2} [(1+\sigma')T_{22(1),1} - \sigma' T_{\gamma\gamma(1),1}]d\xi^2 = 2(1+\sigma')T_{12(1)}. \end{aligned} \tag{A.12}$$

([12]) The method of proof when the displacement is small has been indicated by Goodier. See J. N. Goodier, loc. cit., §4.

Now

$$\int_0^{\xi^1}\int_0^{\xi^2}[(1+\sigma')T_{11(1),22}-\sigma' T_{\gamma\gamma(1),22}]d\xi^1 d\xi^2$$

$$(A.13)\qquad =\int_0^{\xi^1}[(1+\sigma')T_{11(1),2}-\sigma' T_{\gamma\gamma(1),2}]d\xi^1 - \int_0^{\xi^1}[(1+\sigma')T_{11(1),2}-\sigma' T_{\gamma\gamma(1),2}]_{\xi^2=0}d\xi^1.$$

From this and a similar expression derived by interchange of the indices 1 and 2, it is seen that (A.12) takes the form

$$(A.14)\qquad \begin{aligned} u'_{1,2}+u'_{2,1}&+\int_0^{\xi^1}\int_0^{\xi^2}[(1+\sigma')(T_{11(1),22}+T_{22(1),11})-\sigma' T_{\gamma\gamma(1),\beta\beta}]d\xi^1 d\xi^2 \\ &+\int_0^{\xi^1}[(1+\sigma')T_{11(1),2}-\sigma' T_{\gamma\gamma(1),2}]_{\xi^2=0}d\xi^1 \\ &+\int_0^{\xi^2}[(1+\sigma')T_{22(1),1}-\sigma' T_{\gamma\gamma(1),1}]_{\xi^1=0}d\xi^2 = 2(1+\sigma')T_{12(1)}. \end{aligned}$$

If

$$(A.15)\qquad \sigma'=\sigma/(1-\sigma)$$

where σ is Poisson's ratio, then by (A.9) the double integral in (A.14) becomes

$$(A.16)\qquad \begin{aligned} &2\int_0^{\xi^1}\int_0^{\xi^2}(1+\sigma')T_{12(1),12}d\xi^1 d\xi^2 \\ &= 2(1+\sigma')(T_{12(1)}-[T_{12(1)}]_{\xi^1=0}-[T_{12(1)}]_{\xi^2=0}+[T_{12(1)}]_{\xi^1=\xi^2=0}). \end{aligned}$$

Thus (A.14) takes the form

$$(A.17)\qquad u'_{1,2}+u'_{2,1}=G_0(x^0)+G_1(x^0,\xi^1)+G_2(x^0,\xi^2).$$

Since the variables ξ^α have been separated, u'_α can be found. Thus u_α exist satisfying (A.10), the constant σ' having the value $\sigma/(1-\sigma)$ where σ is Poisson's ratio.

Because of (A.1),

$$(A.18)\qquad \int_N u_\alpha T_{\alpha\beta(1),\beta}d\xi^1 d\xi^2 = 0,$$

where, as usual, N denotes a cross-section of the rod (on which x^0 = constant). Since $n_{\beta(0)}$ have the values indicated in (7.24), by Green's theorem (A.18) can be written in the form

$$\oint_N u_\alpha T_{\alpha\beta(1)} n_{\beta(0)} d\tau - \int_N u_{\alpha,\beta} T_{\alpha\beta(1)} d\xi^1 d\xi^2 = 0. \tag{A.19}$$

Because of the boundary condition (A.5) the first integral in (A.19) vanishes. Since $T_{\alpha\beta(1)} = T_{\beta\alpha(1)}$, the remaining integral can be written in the form

$$\int_N \tfrac{1}{2}(u_{\alpha,\beta} + u_{\beta,\alpha}) T_{\alpha\beta(1)} d\xi^1 d\xi^2. \tag{A.20}$$

By substitution from (A.10), it is found that

$$\int_N [2(T_{12(1)})^2 + \tfrac{1}{2}(T_{11(1)} - T_{22(1)})^2 + \tfrac{1}{2}(1 - 2\sigma)(T_{\alpha\alpha(1)})^2] d\xi^1 d\xi^2 = 0. \tag{A.21}$$

Since $1-2\sigma>0$, it is then necessary that

$$T_{\alpha\beta(1)} = 0 \tag{A.22}$$

Thus *Saint-Venant's hypothesis has been proved for the finite displacement of thin rods.*

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICH.

HOMOMORPHISMS AND MODULAR FUNCTIONALS

BY

SAUL GORN

This paper is concerned with complemented modular lattices containing the elements O and I. The first part treats of homomorphisms of the lattice L, their existence, determination and invariant properties. The second considers norms (i.e., sharply positive or, alternatively, strictly monotone modular functionals) and quasi-norms (i.e., positive or monotone modular functionals) on L, their interconnections, and necessary and sufficient conditions for their unicity up to linear transformations.

There are six main parts to the paper, as follows:

1. **The homomorphism theorem.** The dual concepts of σ-ideal and π-ideal are defined for general lattices. Duality is essential throughout the paper. The C-operator, which takes all elements of a subset of L into their complements, is introduced, and C-neutral ideals are defined as those which appear in complementary pairs $\mathfrak{a}$, $\bar{\mathfrak{a}}$. Theorems 1 and 2 state that any C-neutral pair of ideals determine a congruence in L by means of any one of six equivalent conditions. These conditions are recognizable as those appearing in Boolean algebra, but the proof of their equivalence in the general case considered here is far from trivial, since it requires the fundamental Lemma 7. Theorem 3 states that all congruences are thus obtained from C-neutral ideals. Quotient lattices $L/\mathfrak{a}$ are defined, and it is obvious that every homomorph of L is equivalent to an $L/\mathfrak{a}$. For example, consider a regular Carathéodory measure in a metric space, the measure of the space being 1. In the Boolean algebra of measurable sets, the sets of measure 0 and measure 1 are complementary C-neutral ideals, the first σ, the second π, and the quotient lattice is isomorphic with a sublattice of the G_δ's.

2. **The preservation of normal ideals under homomorphism.** The operators c_σ, c_π, and $'$, are defined. By means of the first two we define normal ideals, the upper and lower segments of MacNeille's cuts, whose main reason for existence is to make up for the "gaps" when L is not complete. The main theorem (Theorem 7) states that a homomorphism preserves normality for $\mathfrak{b}\supseteq\mathfrak{a}$; and the pre-image of a normal ideal is normal if $\mathfrak{a}$ is normal. The preliminaries to Theorem 7 state in effect that the operators C, c, and $'$ preserve complementary C-neutrality for pairs of ideals, yielding by iteration at most three pairs from a given $\mathfrak{a}$ and $\bar{\mathfrak{a}}$. It follows that normality in our definition is a proper generalization of Stone's in a Boolean algebra. Lemma 12 gives a connection between neutrality and distributivity parallel to that for complementary neutral elements, a, $\bar{a}$.

Presented to the Society, October 26, 1940; received by the editors December 5, 1940.

3. **Ideal reducibility.** This concept generalizes that of reducibility. Essentially it means that if L were complete, it would be reducible. $L/\mathfrak{a}$ is ideally irreducible if and only if $\mathfrak{a}$ is maximal (Theorem 8—prime, for π-ideals).

4. **Relationship between norms and quasi-norms.** Every quasi-norm determines a complementary pair of C-neutral ideals, $\mathfrak{a}$, $\bar{\mathfrak{a}}$, the elements whose quasi-norms are 0 and 1 (Theorem 9); furthermore (Theorem 10) the quasi-norm determines a norm in $L/\mathfrak{a}$, and every quasi-norm is determined by the $\mathfrak{a}$ and the norm in $L/\mathfrak{a}$.

5. **Unicity of norm.** If L is reducible, its norm is not unique (Lemma 15). This result is elementary and is probably well known. The converse is false; ideal irreducibility is needed. As an example of 4 and 5, since the only irreducible Boolean algebra is $\{O, I\}$, the measure function in 1 is unique if and only if every measurable set is either of measure 0 or of measure 1. The Wilcox-Smiley continuity conditions on a norm are introduced. Essentially, they say that none of the gaps in L affects the norm. If a norm satisfies these conditions, then (Theorems 11 and 12) L is ideally irreducible if and only if the norm is unique, which in turn is true if and only if the complete envelope of L is irreducible. von Neumann's theorem on the unicity of dimension in a continuous geometry is needed in the proof.

6. **Normality.** If a quasi-norm in L satisfies the W.S. conditions, then (Theorem 13) the ideals determined as in 4 are normal. Hence (Theorem 14) the quasi-norm is unique, by 3, if and only if $\mathfrak{a}$ is maximal or prime.

By a *π-ideal* $\mathfrak{a}$ in L, we mean a set of elements such that $a, b \in \mathfrak{a}$ imply $ab \in \mathfrak{a}$, and such that $a \in \mathfrak{a}$, $c \geqq a$ imply $c \in \mathfrak{a}$. The dual of this statement defines a *σ-ideal.* The letters $\mathfrak{a}$ and $\mathfrak{b}$ refer exclusively to ideals.

If $A \subseteq L$, then by CA we mean the set of all complements of all the individual $a \in A$. If $A = \{a\}$, we write Ca instead of CA. Note that $CCA \supseteq A$. By means of Lemma 2 below, it is easy to see that if $\mathfrak{a}$ is a π-ideal, $C\mathfrak{a}$ is the set of all x for which $xa = O$ for some $a \in \mathfrak{a}$; the dual statement holds for σ-ideals.

We call $\mathfrak{a}$ a *C-ideal* if $C\mathfrak{a}$ is a π- or σ-ideal. Note that if $\mathfrak{a}$ is a C-ideal and a π-ideal, then $C\mathfrak{a}$ is a σ-ideal, and vice versa.

We shall never use a $'$ to indicate a complement of an element. On the other hand, a bar over a letter will always indicate that we have before us a uniquely determined complement.

An ideal $\mathfrak{a}$ is called *neutral* (the terminology is due to G. Birkhoff, p. 59, footnote) if $a' \in \mathfrak{a}$ whenever $a \in \mathfrak{a}$ and a' has a common complement with a (i.e., whenever a and a' are perspective, and hence projective); in other words, $\mathfrak{a}$ is neutral if $CC\mathfrak{a} \subseteq \mathfrak{a}$. Thus, $\mathfrak{a}$ is neutral if and only if it is CC-closed, i.e., $CC\mathfrak{a} = \mathfrak{a}$. Notice that in this definition $C\mathfrak{a}$ need not be an ideal. We leave the relationship between neutral and C-ideals an open problem and define a *C-neutral ideal* as a neutral ideal which is also a C-ideal.

It is easy to see that for principal ideals, $\mathfrak{a}(a)$ is neutral if and only if a is neutral, which is the case if and only if $\mathfrak{a}(a)$ is a C-ideal. For principal ideals

the C-ideals and the neutral ideals are the same. Only Lemmas 3 and 6 below are needed to prove these statements.

If a is C-neutral, we write $\bar{a}$ instead of Ca. If a is neutral, we write $\bar{a}$ for its unique complement. By the *complement of a in b*, where $a<b$, we mean the complement of a in $L(O, b)$; dually, by a *complement of a over b*, where $a>b$, we mean a complement of a in $L(b, I)$.

LEMMA 1. *If $a \geqq b \geqq c \geqq d$ and r is a complement of b in $L(a, c)$ and s is a complement of c in $L(a, d)$, then rs is a complement of b in $L(a, d)$. Dually, we get a corresponding theorem using $a \leqq b \leqq c \leqq d$, $r+s$, and $L(c, a)$, $L(d, a)$.*

For, by hypothesis, we have $a \geqq r \geqq c$, $a \geqq s \geqq d$, $b+r=a$, $br=c$, and $s+c=a$, $sc=d$; hence,

$$d = sc = s(rb) = (rs)b,$$

and

$$\begin{aligned} b + rs &= rs + b(s + c) = rs + bs + c = (rs + c) + (bs + c) \\ &= r(s + c) + b(s + c) = r + b = a. \end{aligned}$$

LEMMA 2. *If $xa=O$, then there is a $b \in Ca$ with $b \geqq x$.*

For, let $z \in C(x+a)$; then the dual of Lemma 1 with $O \leqq a \leqq x+a \leqq I$, x for r, and z for s, gives us $x+z \in Ca$.

LEMMA 3. *If $a, a' \in Cb$ and $a' \leqq a$, then $a=a'$.*

For, since $I=a'+b$ and $O=ab$, $a=a(a'+b)=a'+ab=a'$.

LEMMA 4. *If $x \leqq a$ and $b \in Ca$, then there is a $y \in Cx$ for which $y \geqq b$.*

For, since $ab=O$, $xb \leqq ab=O$. It therefore follows, by Lemma 2, that there is a $y \in Cx$ with $y \geqq b$.

LEMMA 5. *If $a_1 \leqq a_2 \leqq a_3$ and $b_1 \geqq b_3$ with $b_i \in Ca_i$, then there is a $b_2 \in Ca_2$ with $b_1 \geqq b_2 \geqq b_3$.*

For, by Lemma 4, we may take $c_2 \leqq b_1$ with $c_2 \in Ca_2$ and $b_2=b_3+a_3c_2 \leqq b_1$. Then

$$a_2 + b_2 = a_2 + a_3c_2 + b_3 = a_3(a_2 + c_2) + b_3 = a_3 + b_3 = I,$$

and

$$a_2b_2 \leqq a_3b_2 = a_3(b_3 + a_3c_2) = a_3b_3 + a_3c_2 = a_3c_2;$$

so that $a_2b_2 \leqq a_3c_2a_2 = O$. Consequently $a_2b_2=O$.

LEMMA 6. *If $b \leqq a$ and a is neutral, then $c \geqq \bar{a}$ for any $c \in Cb$. Dually, if $b \geqq a$ and a is neutral, then $c \leqq \bar{a}$ for any $c \in Cb$.*

For $\bar{a}b \leqq \bar{a}a = O$ and $I=b+c$. Hence $\bar{a}=\bar{a}b+\bar{a}c=\bar{a}c$. In the dual, note that for neutral elements $\bar{a}$ we have $\bar{a}=(\bar{a}+b)(\bar{a}+c)$.

LEMMA 7. *If* $\mathfrak{a}$ *is a C-neutral* π*-ideal, then for any* $x \in L$ *and* $t \in \mathfrak{a}$*, any complement of* tx *in* x *is in* $\bar{\mathfrak{a}}$*. Dually, if* $\mathfrak{a}$ *is a C-neutral* σ*-ideal, then, for any* $x \in L$ *and* $t \in \mathfrak{a}$*, any complement of* $t+x$ *over* x *is in* $\bar{\mathfrak{a}}$*.*

For let u be a complement of tx in x and $y \in Cx$; then an application of Lemma 1 to $O \leqq tx \leqq x \leqq I$ gives $u+y \in C(tx)$. It follows that $O = tx(u+y) = t(u+xy) = tu$. Hence $u \in \bar{\mathfrak{a}}$.

THEOREM 1. *If* $\mathfrak{a}$ *is a C-neutral* π*-ideal and* $\bar{\mathfrak{a}} = C\mathfrak{a}$*, then the following relations between* a *and* b *are equivalent:*

1. *There is a* $t \in \mathfrak{a}$ *with* $ab = (a+b)t$.
1′. *There is a* $u \in \bar{\mathfrak{a}}$ *with* $a+b = ab+u$.
2. *There is a* $t \in \mathfrak{a}$ *with* $at = bt$.
2′. *There is a* $u \in \bar{\mathfrak{a}}$ *with* $a+u = b+u$.
3. *There is a* $t \in \mathfrak{a}$ *and a* $u \in \bar{\mathfrak{a}}$ *with* $a = bt+u$.
3′. *There is a* $t \in \bar{\mathfrak{a}}$ *and a* $u \in \mathfrak{a}$ *with* $a = (b+t)u$.

By duality, we need only consider the proofs that 1′ follows from 1, and that 1 implies 2, which in turn implies 3, which finally implies 1.

In proving that 1′ follows from 1, we may assume that there is a $t \in a$ for which $ab = (a+b)t$, and must find a $u \in \bar{\mathfrak{a}}$ for which $a+b = ab+u$. Using $a+b$ for x in Lemma 7, let u be a complement of $(a+b)t$ in $a+b$; $u \in \bar{\mathfrak{a}}$. Hence $ab+u = a+b$.

Condition 2 follows from 1: $at = a(a+b)t = aab = ab = bab = b(a+b)t = bt$. To show that 2 implies 3, let u be a complement of at in a; $a = bt+u$. But, by Lemma 7 $u \in \bar{\mathfrak{a}}$.

The proof that 3 implies 1 presents the greatest difficulty. We are given a $t \in \mathfrak{a}$ for which $a = bt+u$, where $u \in \bar{\mathfrak{a}}$, and must find an element of $\mathfrak{a}$ which gives ab when multiplied by $a+b$. Since $a = bt+u$, $a+b = b+u$. Applying Lemma 7 to $b+u$ over b yields a $t_1 \in \mathfrak{a}$ with $(b+u)t_1 = b$, i.e., $(a+b)t_1 = b$. Now apply Lemma 7 to $a = bt+u$ over bt; we get a $t_2 \in \mathfrak{a}$ for which $at_2 = (bt+u)t_2 = bt$. Another application of Lemma 7, this time to bt under b, gives us a $u_2 \in \bar{\mathfrak{a}}$ with $at_2 + u_2 = bt + u_2 = b$. It follows that $a+b = a+u_2$. This permits us a final application of Lemma 7, to $a+u_2$ over a, providing a $t_3 \in \mathfrak{a}$ for which $(a+b)t_3 = (a+u_2)t_3 = a$. Using this last equation with the corresponding expression for b (from our first use of Lemma 7), we get $(a+b)t_1t_3 = (a+b)t_1(a+b)t_3 = ba$, where $t_1t_3 \in \mathfrak{a}$ since $t_1, t_3 \in \mathfrak{a}$.

THEOREM 2. *Define* $a \equiv b$ *to mean that any one of the six equivalent conditions of Theorem* 1 *holds between* a *and* b*. Then* $\equiv$ *is a congruence relation, the elements* $\equiv I$ *form* $\mathfrak{a}$*, and the elements* $\equiv O$ *form* $\bar{\mathfrak{a}}$*.*

For, since $aI = aI$, $a \equiv a$; if $at = bt$, then $bt = at$, so that $a \equiv b$ implies $b \equiv a$. If $a \equiv b \equiv c$, then $at_1 = bt_1$ and $bt_2 = ct_2$, where $t_i \in \mathfrak{a}$; it follows that $t_1t_2 \in \mathfrak{a}$ and $at_1t_2 = ct_1t_2$. Hence $a \equiv c$.

Suppose $a \equiv b$ and $c \equiv d$. On the one hand $at_1 = bt_1$, $ct_2 = dt_2$, where $t_i \in \mathfrak{a}$; $t_1t_2 \in \mathfrak{a}$ and $act_1t_2 = bdt_1t_2$, hence $ac \equiv bd$. On the other hand, $a + u_1 = b + u_1$, $c + u_2 = d + u_2$, where $u_i \in \bar{\mathfrak{a}}$; $u_1 + u_2 \in \bar{\mathfrak{a}}$ and $(a+c)+(u_1+u_2) = (b+d)+(u_1+u_2)$, hence $a + c \equiv b + d$.

Finally, $a \equiv I$, $at = It$ for some $t \in \mathfrak{a}$, $at = t$ for some $t \in \mathfrak{a}$, $a \geqq t$ for some $t \in \mathfrak{a}$, and $a \in \mathfrak{a}$ are equivalent conditions.

THEOREM 3. *If θ is a congruence in L, then the set of elements $\equiv I$ is a C-neutral π-ideal $\mathfrak{a}$, and the set of elements $\equiv O$ is the C-neutral σ-ideal $\bar{\mathfrak{a}}$, and θ is the same as the congruence determined by $\mathfrak{a}$ (or $\bar{\mathfrak{a}}$) in Theorem* 2. (*This is an amplification of Birkhoff's Theorem* 4.3.)

If $a \equiv I$ and $b \geqq a$, then $b \equiv I$; if $a \equiv I$ and $b \equiv I$, then $ab \equiv I$; hence the elements $\equiv I$ form a π-ideal, $\mathfrak{a}$; the dual statement holds for the elements $\equiv O$. But if $a \equiv I$ and $ab = O$, then $O = ab \equiv Ib = b$; hence the elements $\equiv O$ form the set $C\mathfrak{a}$, which is therefore an ideal, and $\mathfrak{a}$ is a C-ideal; but dually we have $\mathfrak{a} = CC\mathfrak{a}$, so that $\mathfrak{a}$ is neutral. Now if $a \equiv b(\theta)$, then $ab \equiv a + b(\theta)$; take t a complement of ab in $a+b$, so that $ab + t = a + b$, and $abt = O$; condition 1′ of Theorem 1 then follows, since $t = (a+b)t \equiv abt = O$. Hence $a \equiv b(\mathfrak{a})$. If $a \equiv b(\mathfrak{a})$, then there is a $u \in \mathfrak{a}$ for which $a + u = b + u$; $u \equiv O(\theta)$, hence $a \equiv a + u = b + u \equiv b(\theta)$.

COROLLARY. *If L is a complemented modular lattice, then the congruences in L are in* (1, 1) *correspondence with the C-neutral π-ideals (σ-ideals) in L.*

The proof is obvious.

If $\mathfrak{a}$ is C-neutral, we define A_a to mean the set of all $x \equiv a(\mathfrak{a})$; we also define $L/\mathfrak{a}$ to mean the set of all A_a, where $A_a + A_b = A_{a+b}$, $A_aA_b = A_{ab}$, and hence $A_a \leqq A_b$ if and only if $ab \equiv a(\mathfrak{a})$. Note that $L/\mathfrak{a}$ is a complemented, modular lattice with A_O and A_I for O and I. A^* is defined to mean the set of all A_a for which $a \in A$.

COROLLARY. *If L is a complemented, modular lattice and is homomorphic to the lattice L', then L' has an O and I, the elements of L homomorphic to O form a C-neutral σ-ideal $\mathfrak{a}$, those homomorphic to I form $\bar{\mathfrak{a}}$, and L' is isomorphic to $L/\mathfrak{a}$.*

The proof is obvious, since "a and b have the same homomorph" is a congruence relation.

LEMMA 8. *If $d^* \in Ca^*$, then there is a $b \in Ca$ for which $b^* = d^*$. The converse is obvious.*

For if the congruence ideals are $\mathfrak{a}$, $\bar{\mathfrak{a}}$, the first being a σ-ideal, then $(ad)^* = a^*d^* = O^*$, i.e., $ad \in \mathfrak{a}$. Take $t \in Cad$, so that $t \in \bar{\mathfrak{a}}$; then $dt \equiv d$, and $a(dt) = (ad)t = O$. By Lemma 2 there is a $b \geqq dt$ with $b \in Ca$; therefore $b = b + dt \equiv b + d$ and consequently $b^* = (b+d)^* = b^* + d^*$. Thus $b^* \geqq d^*$ and $b^*, d^* \in Ca^*$. We may now apply Lemma 3 in $L^* = L/\mathfrak{a}$ to give $b^* = d^*$.

Theorem 4. *There is a* (1, 1) *correspondence between the ideals of* L, $\mathfrak{b}\supseteq\mathfrak{a}$, *and the ideals of* L^*, $\mathfrak{B}$, *given by*

$$\mathfrak{B} = \mathfrak{b}^*, \quad \mathfrak{b} = \bigcup_{b^*\in\mathfrak{B}} b^*. \tag{1}$$

Thus, if $\mathfrak{b}_1^*=\mathfrak{b}_2^*$ *and* $\mathfrak{b}_i\supseteq\mathfrak{a}$, *then* $\mathfrak{b}_1=\mathfrak{b}_2$. *This correspondence is also* (1, 1) *for the C-neutral ideals.*

In our proof we may restrict ourselves to π-ideals. Let the ideal $\mathfrak{b}\supseteq\mathfrak{a}$, $b\in\mathfrak{b}$, and $d\equiv b$; then $dt=bt$ for some $t\in\mathfrak{a}\subseteq\mathfrak{b}$. Hence $d\geqq dt=bt\in\mathfrak{b}$, so that $d\in\mathfrak{b}$. Thus any ideal $\supseteq\mathfrak{a}$ contains complete congruence classes, and (1) follows. If $a^*\geqq b^*$ and $b^*\in\mathfrak{b}^*$, then $(ab)^*=a^*b^*=b^*$, and $ab\equiv b$; there is consequently a $t\in\mathfrak{a}$ with $abt=bt$. It therefore follows that $a\geqq bt\in\mathfrak{b}$, $a\in\mathfrak{b}$, and $a^*\in\mathfrak{b}^*$. If a^*, $b^*\in\mathfrak{b}^*$, then $ab\in\mathfrak{b}$, and $a^*b^*=(ab)^*\in\mathfrak{b}^*$. Thus $\mathfrak{b}^*$ is an ideal. If $\mathfrak{B}$ is an ideal in L^* and (1) holds, then $\mathfrak{b}$ is an ideal $\supseteq\mathfrak{a}=I^*$, since $I^*\in\mathfrak{B}$, and the typical argument holds, namely: if b_1, $b_2\in\mathfrak{b}$, then $b_i^*\in\mathfrak{B}$ so that $(b_1b_2)^*\in\mathfrak{B}$, and hence $b_1b_2\in\mathfrak{b}$; if $b_1\geqq b_2\in\mathfrak{b}$, then $b_1^*\geqq b_2^*\in\mathfrak{B}$, $b_1^*\in\mathfrak{B}$, and consequently $b_1\in\mathfrak{b}$.

Suppose $\mathfrak{b}$ is C-neutral. Then $C\mathfrak{b}^*=$ the set of $Cb^*=$ the image of the Cb (by Lemma 8) $=(C\mathfrak{b})^*$; since $C\mathfrak{b}$ is a σ-ideal, so is $C\mathfrak{b}^*$ by the dual of the preceding; it follows that $\mathfrak{b}^*$ is a C-ideal.

The dual argument yields $CC\mathfrak{b}^*=C(C\mathfrak{b})^*=(CC\mathfrak{b})^*=\mathfrak{b}^*$; $\mathfrak{b}^*$ is C-neutral.

The same reasoning gives: if $\mathfrak{b}^*$ is C-neutral, $(C\mathfrak{b})^*$ is an ideal, and $\mathfrak{b}^*=CC\mathfrak{b}^*=(CC\mathfrak{b})^*$; hence to prove $\mathfrak{b}$ C-neutral, we have only to show that the pre-images of $(C\mathfrak{b})^*$ and $(CC\mathfrak{b})^*$ are $C\mathfrak{b}$ and $CC\mathfrak{b}$; i.e., that if $x\in C\mathfrak{b}$ and $y\equiv x$, then $y\in C\mathfrak{b}$, and if $x\in CC\mathfrak{b}$ and $y\equiv x$, then $y\in CC\mathfrak{b}$. If $x\in C\mathfrak{b}$ and $y\equiv x$, then $y^*=x^*\in C\mathfrak{b}^*$. Lemma 8 then gives us an $a\in Cy$ for which $a\equiv b$. Thus $y\in Ca$ and $a\in\mathfrak{b}$. Hence $y\in C\mathfrak{b}$ and $C\mathfrak{b}$ is an ideal. The dual argument makes $CC\mathfrak{b}$ an ideal. Now $CC\mathfrak{b}\supseteq\mathfrak{a}$, so that, since $\mathfrak{b}^*=(CC\mathfrak{b})^*$, $CC\mathfrak{b}=\mathfrak{b}$.

If $A\subseteq L$, we define $c_\sigma A$ to mean the set of all $x\in L\leqq$ every $a\in A$, and $c_\pi A$ to mean the set of all $x\in L\geqq$ every $a\in A$. A σ-ideal (π-ideal) is *normal*[1] if $\mathfrak{a}=c_\sigma c_\pi\mathfrak{a}$ ($\mathfrak{a}=c_\pi c_\sigma\mathfrak{a}$).

The $c_\sigma A$, $c_\pi A$ are normal. The normal ideals of either type form the complete envelope of L, being the lower and upper segments of MacNeille's cuts.

Lemma 9. $a=$g.l.b. a^* *if and only if* $a\in c_\sigma\bar{\mathfrak{a}}$; $a=$l.u.b. a^* *if and only if* $a\in c_\pi\mathfrak{a}$, *where* $\mathfrak{a}$ *is a C-neutral* σ*-ideal. If* $b\in Ca$, *then* $a=$g.l.b. a^* *if and only if* $b=$l.u.b. b^*.

The proof divides naturally into three parts:

1. If $a\in c_\sigma\bar{\mathfrak{a}}$ and $b\equiv a$, then there is a $t\in\bar{\mathfrak{a}}$ for which $at=bt$; since $a\leqq t$, it follows that $a=bt$ and $a\leqq b$.

---

(1) For the introduction of these ideas see the author's abstract 45-1-17; Wilcox and Smiley's *Metric lattices* in Annals of Mathematics, (2), vol. 40 (1939), and §33 in Birkhoff.

2. If $a \equiv b$ implies $a \leqq b$, then for any $t \in \bar{\mathfrak{a}}$ we have, as a consequence of $a \equiv at$, $a \leqq at$; thus $a = at$, and $a \leqq t$.

3. Suppose $a = \text{g.l.b. } a^*$ and $b \in C\mathfrak{a}$, but $b \neq \text{l.u.b. } b^*$, so that part 2 permits us to say that there is a $u \in \mathfrak{a}$ with $b + u > u$. By Lemmas 3 and 4 there is an $a_1 < a$ with $b + u \in C a_1$. It follows that $a_1 b = O$ and $a_1 + b \equiv a_1 + b + u = I$; hence Lemma 8 yields an $a_2 \in Cb$ for which $a_2 \equiv a_1$, say $a_2 t = a_1 t$ with $t \in \bar{\mathfrak{a}}$. Then $a_2 t \leqq a_1 < a$. But

$$a = a(b + a_2) \equiv a(b + a_2 t) = ab + a_2 t = a_2 t.$$

This contradicts the fact that $a = \text{g.l.b. } a^*$.

The proof is then completed by appealing to duality.

Corollary. *$Cc_\sigma = c_\pi C$, $Cc_\pi = c_\sigma C$. In other words, for any C-neutral σ-ideal* $\mathfrak{a}$ *we have*

$$Cc_\sigma \bar{\mathfrak{a}} = c_\pi \mathfrak{a}, \qquad Cc_\pi \mathfrak{a} = c_\sigma C\mathfrak{a} = c_\sigma \bar{\mathfrak{a}}.$$

For Lemma 9 gives us: if $b \in C\mathfrak{a}$, then $a \in c_\sigma \bar{\mathfrak{a}}$ if and only if $b \in c_\pi \mathfrak{a}$. Hence $c_\sigma \bar{\mathfrak{a}}$ and $c_\pi \mathfrak{a}$ are complementary C-ideals.

Theorem 5. *If* $\mathfrak{a}$ *is a C-neutral σ-ideal, then* $c_\sigma \bar{\mathfrak{a}}$ *and* $c_\pi \mathfrak{a}$ *are complementary C-neutral ideals, and* $c_\pi c_\sigma \bar{\mathfrak{a}}$ *and* $c_\sigma c_\pi \mathfrak{a}$ *are complementary C-neutral normal ideals containing* $\bar{\mathfrak{a}}$ *and* $\mathfrak{a}$, *respectively.*

For $c_\sigma \bar{\mathfrak{a}}$, $c_\pi \mathfrak{a}$ are C-ideals, and

$$CCc_\pi \mathfrak{a} = Cc_\sigma \bar{\mathfrak{a}} = c_\pi C\bar{\mathfrak{a}} = c_\pi \mathfrak{a};$$

dually for $c_\sigma \bar{\mathfrak{a}}$; hence $c_\pi \mathfrak{a}$ and $c_\sigma \bar{\mathfrak{a}}$ are complementary C-neutral ideals for any C-neutral $\mathfrak{a}$, $\bar{\mathfrak{a}}$. Applying this to $c_\pi \mathfrak{a}$ and $c_\sigma \bar{\mathfrak{a}}$ instead of $\bar{\mathfrak{a}}$ and $\mathfrak{a}$, we get the remaining parts of the theorem.

With Stone we define: *if $A \subseteq L$, then by A' we mean the set of all $b \in L$ for which*

Case 1. *$ab = O$ for all $a \in A$ if A is a σ-ideal.*

Case 2. *$a + b = I$ for all $a \in A$ if A is a π-ideal.*

Note that this operation is a polarity (see Birkhoff §32), so that, similarly to c_σ and c_π, we have: If $A \subseteq B$, then

$$A' \supseteq B', \qquad A \subseteq A'', \qquad A''' = A'.$$

Theorem 6. *If* $\mathfrak{a}$ *is a C-neutral σ-ideal, then*

$$\mathfrak{a}' = c_\sigma \bar{\mathfrak{a}} = (c_\sigma \bar{\mathfrak{a}})'' = (c_\sigma c_\pi \mathfrak{a})'$$

and

$$\mathfrak{a}'' = (c_\sigma \bar{\mathfrak{a}})' = (c_\sigma c_\pi \mathfrak{a})'' = c_\sigma c_\pi \mathfrak{a}.$$

We will prove part of the theorem directly and will obtain the rest by continued application of the polarity properties.

1. If $v \in c_\sigma \bar{\mathfrak{a}}$ and $u \in c_\sigma c_\tau \mathfrak{a}$, take $w \in Cv$. Then $w \in c_\tau \mathfrak{a}$, so that $u \leqq w$; therefore $uv \leqq wv = O$, and $v \in (c_\sigma c_\tau \mathfrak{a})'$, $u \in (c_\sigma \bar{\mathfrak{a}})'$; i.e.,

$$c_\sigma \bar{\mathfrak{a}} \subseteq (c_\sigma c_\tau \mathfrak{a})', \qquad c_\sigma c_\tau \mathfrak{a} \subseteq (c_\sigma \bar{\mathfrak{a}})'.$$

2. If $v \in \mathfrak{a}$, let $t \in \bar{\mathfrak{a}}$ and let u be a complement of vt in v; by Lemma 7 we have $u \in \mathfrak{a}$ and $v = vt + u$. Now $uv = O$, therefore $u = uv = O$, $v = vt$, and hence $v \leqq t$. It therefore follows that $v \in c_\sigma \bar{\mathfrak{a}}$ and $\mathfrak{a}' \subseteq c_\sigma \bar{\mathfrak{a}}$.

3. We now have $\mathfrak{a}' \subseteq c_\sigma \bar{\mathfrak{a}} \subseteq (c_\sigma c_\tau \mathfrak{a})'$ and $c_\sigma c_\tau \mathfrak{a} \subseteq (c_\sigma \bar{\mathfrak{a}})'$; the last yields $(c_\sigma \bar{\mathfrak{a}})'' \subseteq (c_\sigma c_\tau \mathfrak{a})'$, hence

$$\mathfrak{a}' \subseteq c_\sigma \bar{\mathfrak{a}} \subseteq (c_\sigma \bar{\mathfrak{a}})'' \subseteq (c_\sigma c_\tau \mathfrak{a})';$$

therefore

$$\mathfrak{a}'' \supseteq (c_\sigma \bar{\mathfrak{a}})' \supseteq (c_\sigma c_\tau \mathfrak{a})'' \supseteq c_\sigma c_\tau \mathfrak{a}.$$

Now use $c_\sigma \bar{\mathfrak{a}}$ in the preceding instead of $\mathfrak{a}$:

$$(c_\sigma \bar{\mathfrak{a}})' \subseteq c_\sigma c_\tau \mathfrak{a} \subseteq (c_\sigma c_\tau \mathfrak{a})'' \subseteq (c_\sigma \bar{\mathfrak{a}})',$$

$$(c_\sigma \bar{\mathfrak{a}})'' \supseteq (c_\sigma c_\tau \mathfrak{a})' \supseteq (c_\sigma \bar{\mathfrak{a}})'' \supseteq c_\sigma \bar{\mathfrak{a}}.$$

If we match the last four statements in pairs, we get

$$\mathfrak{a}'' \supseteq c_\sigma c_\tau \mathfrak{a} = (c_\sigma c_\tau \mathfrak{a})'' = (c_\sigma \bar{\mathfrak{a}})',$$

$$\mathfrak{a}' \subseteq c_\sigma \bar{\mathfrak{a}} \subseteq (c_\sigma \bar{\mathfrak{a}})'' = (c_\sigma c_\tau \mathfrak{a})'.$$

But $\mathfrak{a} \subseteq c_\sigma c_\tau \mathfrak{a}$; hence $\mathfrak{a}' \supseteq (c_\sigma c_\tau \mathfrak{a})'$ and $\mathfrak{a}'' \subseteq (c_\sigma c_\tau \mathfrak{a})''$, giving the theorem.

COROLLARY. *A C-neutral ideal is normal if and only if* $\mathfrak{a} = \mathfrak{a}''$.

The proof is obvious.

In a Boolean algebra all elements are neutral, so that $CCA = A$ for all $A \subseteq L$; if $\mathfrak{a}$ is a σ-ideal and $x, y \in C\mathfrak{a}$, then $\bar{x} + \bar{y} \in \mathfrak{a}$, and $xy \in C\mathfrak{a}$, and any ideal is C-neutral. Thus our definition of normality is a proper generalization of Stone's ($\mathfrak{a}'' = \mathfrak{a}$) in a Boolean algebra. Henceforth we may write $\mathfrak{a}'$ for $c_\sigma \bar{\mathfrak{a}}$, $\mathfrak{a}''$ for $c_\sigma c_\tau \mathfrak{a}$, $\bar{\mathfrak{a}}'$ for $c_\tau \mathfrak{a}$, and $\bar{\mathfrak{a}}''$ for $c_\tau c_\sigma \bar{\mathfrak{a}}$.

LEMMA 10. *If* $b_1 \equiv b_2(\mathfrak{a})$ *and* $b_1 \equiv b_2(\mathfrak{a}')$, *then* $b_1 = b_2$. *Similarly for* $\mathfrak{a}'$ *and* $\mathfrak{a}''$.

For take a_3 complementary to $b_1 b_2$ in $b_1 + b_2$: $a_3 b_1 b_2 = O$, $a_3 + b_1 b_2 = b_1 + b_2$. Now $b_1 b_2 \equiv b_i \equiv b_1 + b_2$ by either ideal, hence $a_3 = a_3(b_1 + b_2) \equiv a_3 b_1 b_2 = O$ by either ideal, i.e., $a_3 \in \mathfrak{a} \cap \mathfrak{a}'$, so that $a_3 = O$. Thus $b_1 b_2 \leqq b_i \leqq b_1 + b_2 = b_1 b_2$, hence $b_1 = b_1 b_2 = b_2$.

LEMMA 11. *If* $a' \in \mathfrak{a}'$, *then* $x \equiv a'$ *if and only if* $x = a + a'$ *with* $a \in \mathfrak{a}$; $a' + a_1 = a' + a_2$ *if and only if* $a_1 = a_2$, *so that* $(a')^*$ *is isomorphic with* $\mathfrak{a}$ *by* $a' + a \longleftrightarrow a$.

For there is a $t \in \bar{\mathfrak{a}}$ with $xt = a't = a'$. Let a be a complement of xt in x.

By Lemma 7 $a\in\mathfrak{a}$ and hence $x=xt+a=a'+a$. If $a'+a_1=a'+a_2$ with $a_i\in\mathfrak{a}$, then $a_1+a_2\in\mathfrak{a}$, and therefore

$$\begin{aligned} a_1 &= a_1 + a'(a_1 + a_2) = (a_1 + a')(a_1 + a_2) = (a_2 + a')(a_1 + a_2) \\ &= a_2 + a'(a_1 + a_2) = a_2. \end{aligned}$$

LEMMA 12. *For any $a\in\mathfrak{a}$, $a'\in\mathfrak{a}'$ and $x\in L$ we have $x(a+a')=xa+xa'$.*

For $x(a+a')\equiv xa'\in\mathfrak{a}'$; hence Lemma 11 yields an $a_1\in\mathfrak{a}$ for which $x(a+a')=xa'+a_1$. Thus $xa\equiv a_1(\mathfrak{a}')$. But $xa\equiv a_1(\mathfrak{a})$, since both are in $\mathfrak{a}$. Consequently, by Lemma 10, $a_1=xa$.

We now define $A+B$ to mean the set of all $a+b$ with $a\in A$, $b\in B$. Similarly for AB.

LEMMA 13. *If $\mathfrak{b}$ is a C-neutral ideal $\supseteq\mathfrak{a}$, then $\mathfrak{a}+\mathfrak{b}'$ is a C-neutral σ-ideal with $\mathfrak{a}\mathfrak{b}'$ as its complement, and $(\mathfrak{a}+\mathfrak{b}')'=\mathfrak{a}'\cap\mathfrak{b}''$.*

For $\mathfrak{b}'\subseteq\mathfrak{a}'$ and $(a_1+b_1')+(a_2+b_2')=(a_1+a_2)+(b_1'+b_2')$; if $x\leqq a+b'$, since $b'\in\mathfrak{b}'\subseteq\mathfrak{a}'$, Lemma 12 gives $x=x(a+b')=xa+xb'$, with $xa\in\mathfrak{a}$, $xb'\in\mathfrak{b}'$. Thus $\mathfrak{a}+\mathfrak{b}'$ is an ideal. Now let $x\in C\mathfrak{a}+\mathfrak{b}'$. Since $a+b'\equiv b'\in\mathfrak{b}'\subseteq\mathfrak{a}'$, and since, by Lemma 4, we may take $x\leqq t'\in C\mathfrak{b}'$, it follows that $t'\in\bar{\mathfrak{b}}'$ and $x\equiv t'$. Thus there is an $a_1\in\mathfrak{a}$ for which $x+a_1=t'+a_1=t'$, since $\bar{\mathfrak{b}}'\subseteq\bar{\mathfrak{a}}'$. But $x+a_1$ over x has a complement $t\in\bar{\mathfrak{a}}$ by Lemma 7; hence $x=tt'\in\bar{\mathfrak{a}}\bar{\mathfrak{b}}'$. It follows that $C(\bar{\mathfrak{a}}\bar{\mathfrak{b}}')\subseteq\mathfrak{a}+\mathfrak{b}'$, and, dually that $C(\mathfrak{a}+\mathfrak{b}')\subseteq\bar{\mathfrak{a}}\bar{\mathfrak{b}}'$, so that $CC(\mathfrak{a}+\mathfrak{b}')\subseteq\mathfrak{a}+\mathfrak{b}'$, and dually for $\bar{\mathfrak{a}}\bar{\mathfrak{b}}'$; thus $\bar{\mathfrak{a}}\bar{\mathfrak{b}}'=CC\bar{\mathfrak{a}}\bar{\mathfrak{b}}'\subseteq C(\mathfrak{a}+\mathfrak{b}')\subseteq\bar{\mathfrak{a}}\bar{\mathfrak{b}}'$, and the first statement follows.

If $x\in\mathfrak{a}'\cap\mathfrak{b}''$, then $xa=O$ and $xb'=O$ for all a, b'; by Lemma 12, $x(a+b')=O$, and $x\in(\mathfrak{a}+\mathfrak{b}')'$.

Finally, if $x(a+b')=O$, then $xa=xb'=O$, and $x\in\mathfrak{a}'\cap\mathfrak{b}''$, q. e. d.

COROLLARY. *If $\mathfrak{b}$ is a C-neutral ideal $\supseteq\mathfrak{a}$, then $(\mathfrak{b}'')^*=(\mathfrak{a}+\mathfrak{b}')^{*\prime}$.*

For if $x^*\in(\mathfrak{b}'')^*$, then $x\in\mathfrak{b}''$ by Theorem 4, since $\mathfrak{b}''\supseteq\mathfrak{a}''\supseteq\mathfrak{a}$. Therefore $xb'=O$ for all $b'\in\mathfrak{b}'$. Hence $x(a+b')\equiv O$ for all $a\in\mathfrak{a}$ and $b'\in\mathfrak{b}'$. Consequently,

$$x^*(a+b')^*=(x(a+b'))^*=O^*$$

and hence $x^*\in(\mathfrak{a}+\mathfrak{b}')^{*\prime}$.

If $x^*\in(\mathfrak{a}+\mathfrak{b}')^{*\prime}$, then as before, $x(a+b')\equiv O$, and $xb'\equiv O$ for all $b'\in\mathfrak{b}'$; but $xb'\in\mathfrak{b}'\subseteq\mathfrak{a}'$ and Lemma 10 gives $xb'=O$. Hence $x\in\mathfrak{b}''$, and $x^*\in(\mathfrak{b}'')^*$.

LEMMA 14. *If $\mathfrak{a}$ is normal and $bx\in\mathfrak{a}$ for all $b\in\mathfrak{b}$, then $b''x\in\mathfrak{a}$ for all $b''\in\mathfrak{b}''$.*

For $a'bx=O$ for all $a'\in\mathfrak{a}'$ and $b\in\mathfrak{b}$; hence $a'x\in\mathfrak{b}'$ for all $a'\in\mathfrak{a}'$; therefore $b''a'x=O$ for all $b''\in\mathfrak{b}''$ and $a'\in\mathfrak{a}'$, so that $b''x\in\mathfrak{a}''=\mathfrak{a}$ for all $b''\in\mathfrak{b}''$.

COROLLARY. *If $\mathfrak{a}$ is normal and $\mathfrak{b}$ is a C-neutral ideal $\supseteq\mathfrak{a}$, then $\mathfrak{b}^{*\prime}=\mathfrak{b}''^{*\prime}$.*

The proof is obvious.

THEOREM 7. *If* $\mathfrak{a}$ *is a C-neutral ideal and* $\mathfrak{b}$ *is a C-neutral ideal* $\supseteq \mathfrak{a}$, *then*:

1. *If* $\mathfrak{b}$ *is normal, so is* $\mathfrak{b}^*$.
2. *If* $\mathfrak{a}$ *is normal and* $\mathfrak{b}^*$ *is normal, so is* $\mathfrak{b}$.
3. *If* $\mathfrak{a}$ *is normal, the correspondence of Theorem* 4 *preserves normality in both directions.*

For 1. By the corollary to Lemma 13, if $\mathfrak{b}$ is normal, then $\mathfrak{b}^* = (\mathfrak{b}'')^* = (\mathfrak{a}+\mathfrak{b}')^{*\prime}$, so that $\mathfrak{b}^*$ is normal.

2. By the last corollary, if $\mathfrak{a}$ is normal and $\mathfrak{b}^*$ is normal, then, using the first part of the proof, $(\mathfrak{b}'')^* = (\mathfrak{b}'')^{*\prime\prime} = (\mathfrak{b}^*)'' = \mathfrak{b}^*$. Hence, by Theorem 4, $\mathfrak{b}'' = \mathfrak{b}$.

3. The proof is now obvious.

We note that it is not hard to show, using Lemma 3 in $L(a, I)$ and Lemma 9, that a^* can have a g.l.b. a and an l.u.b. if and only if $\mathfrak{a}$ is principal; and that $L/\mathfrak{a}$ is isomorphic with $\mathfrak{a}'$ if and only if $\mathfrak{a}'$, and hence $\mathfrak{a}$, is principal. It also follows (using $\mathfrak{b} = \mathfrak{a}$ in Lemma 13) that

$$(\mathfrak{a} + \mathfrak{a}')' = \{O\}, \qquad (\bar{\mathfrak{a}}\bar{\mathfrak{a}}')' = \{I\}, \qquad (\mathfrak{a} + \mathfrak{a}')'' = L, \qquad (\bar{\mathfrak{a}}\bar{\mathfrak{a}}')'' = L,$$

so that $\mathfrak{a}+\mathfrak{a}'$ is normal if and only if $\mathfrak{a}+\mathfrak{a}' = L$, which is possible if and only if $\mathfrak{a}$ is principal (for $\mathfrak{a}' \cap \mathfrak{a}'' = \{O\}$). Thus $\mathfrak{a}+\mathfrak{a}'$ is not normal if $\mathfrak{a}$ is not principal, even if $\mathfrak{a}$ is normal. In Stone's terminology, $\mathfrak{a}+\mathfrak{a}'$ is a *barrier ideal* in such a case. Stone has pointed out that the homomorphism $L \to L^*$ need not preserve normal ideals (we have needed $\mathfrak{b} \supseteq \mathfrak{a}$ to obtain such a correspondence). Indeed, using $\mathfrak{b} = \mathfrak{a}$ in the corollary to Lemma 13 we see that, if $\mathfrak{a}$ is normal, $(\mathfrak{a}+\mathfrak{a}')^{*\prime} = (\mathfrak{a}'')^* = \mathfrak{a}^* = O^*$, so that $(\mathfrak{a}')^*$ is not normal if $\mathfrak{a}$ is not principal (we have already seen that $\mathfrak{a}+\mathfrak{a}'$ is not normal, so that we are not contradicting Theorem 7), for $(\mathfrak{a}+\mathfrak{a}')^* = (\mathfrak{a}')^*$, and hence $(\mathfrak{a}')^{*\prime\prime} = L^*$; but $(\mathfrak{a}')^* = L$ only if $\mathfrak{a}+\mathfrak{a}' = L$.

L is called *ideally irreducible* if it contains no normal C-neutral σ-ideal (π-ideal) $\neq O^\sigma$ or I^σ (O^π or I^π). Every ideally irreducible L is also irreducible. If L is complete, the two concepts coincide.

THEOREM 8. *If* $\mathfrak{a}$ *is a normal C-neutral* σ*-ideal* (π*-ideal*) $\neq O^\sigma, I^\sigma$ (O^π, I^π), *then* $L/\mathfrak{a}$ *is ideally irreducible if and only if* $\mathfrak{a}$ *is maximal* (*prime*) *in the partially ordered set of normal C-neutral* σ*-ideals* (π*-ideals*).

This is an obvious consequence of Theorem 7.

By a *quasi-norm* on L, $r(x)$, we mean a positive (or monotone) modular functional. If it is sharply positive, we call it a *norm*. It is called *normalized* if $r(O) = 0$, $r(I) = 1$.

THEOREM 9. *If* $r(x)$ *is a normalized quasi-norm, then the set of* $x \in L$ *for which* $r(x) = 1$ *is a C-neutral* π*-ideal* $\bar{\mathfrak{a}}$, *and* $\mathfrak{a}$ *is the set of all* $x \in L$ *for which* $r(x) = 0$; *furthermore,* $a \equiv b(\mathfrak{a})$ *if and only if* $r(ab) = r(a+b)$. (*This is an extension of Lemma* 1 *in* §87, *Birkhoff*.)

The proof falls into seven sections.

1. If $x \geqq y$ and $r(y)=1$, then $r(x) \geqq r(y)=1$, so that $r(x)=1$. If $r(x)=r(y)=1$, then $r(x+y)=1$, so that

$$r(xy) = r(x) + r(y) - r(x+y) = 1.$$

Thus $\bar{\mathfrak{a}}$ is a π-ideal.

2. If $y \in Cx$, then

$$1 = r(I) = r(x+y) = r(x) + r(y) - r(xy) = r(x) + r(y).$$

3. If $y, y' \in Cx$, then $r(y) = 1 - r(x) = r(y')$.

4. If $x \in \bar{\mathfrak{a}}$ and $x' \in CCx$, then, $r(x') = r(x) = 1$; $CC\mathfrak{a} \subseteq \mathfrak{a}$.

5. If $x \in Cy$, then $x \in \bar{\mathfrak{a}}$ if and only if $r(y)=0$, by 2. Since, by the dual of 1, the latter class is a σ-ideal, $C\bar{\mathfrak{a}}$ is a σ-ideal.

6. By 4 and 5, $\bar{\mathfrak{a}}$ is C-neutral, and $\mathfrak{a}$ is the set of x with $r(x)=0$.

7. If $a \equiv b(\mathfrak{a})$, then $a+b = ab+u$ where $u \in \mathfrak{a}$. But $r(u) = r(abu) = 0$. Therefore $r(a+b) = r(ab+u) = r(ab) + r(u) - r(abu) = r(ab)$. If $r(ab) = r(a+b)$, take u a complement of ab in $a+b$: $a+b = ab+u$, $abu = O$; then

$$r(a+b) = r(ab) + r(u) - r(abu) = r(ab) + r(u)$$

so that $r(u)=0$, $u \in \mathfrak{a}$, and $a+b \equiv ab(\mathfrak{a})$.

COROLLARY. *If $a \equiv b(\mathfrak{a})$, then $r(a)=r(b)$. If $a<b$ and $r(a)=r(b)$, then $a \equiv b(\mathfrak{a})$.*

To prove the first part, suppose $a \equiv b(\mathfrak{a})$. Then

$$ab \leqq \begin{Bmatrix} a \\ b \end{Bmatrix} \leqq a+b$$

yields

$$r(ab) \leqq \begin{Bmatrix} r(a) \\ r(b) \end{Bmatrix} \leqq r(a+b) = r(ab);$$

hence $r(a) = r(ab) = r(b)$.

As for the second, let u be a complement of a in b: $b = a+u$, $au = O$; then $r(a) = r(b) = r(a) + r(u)$, $r(u) = 0$, $u \in \mathfrak{a}$, and $a \equiv b(\mathfrak{a})$.

THEOREM 10. *If $r(x)$ is a quasi-norm in L, then $r(x^*) = r(x)$ is a norm in $L^* = L/\mathfrak{a}$, and every norm in L^* determines a quasi-norm in L by $r(x) = r(x^*)$. There is a (1, 1) correspondence between the norms of L^* and the quasi-norms of L which determine $\mathfrak{a}$.*

The result is obvious except for the fact that $r(x^*)$ is sharply positive: if $x^* < y$, then $x^* y^* = x^* \neq y^*$, so that $xy \equiv x$ and $x \not\equiv y$; there is then a $t \in \mathfrak{a}$ for which $xyt = xt$, $xt < y$, and hence $xt \not\equiv y$. By the above corollary we must then have $r(x) = r(xt) < r(y)$, and consequently $r(x^*) < r(y^*)$.

Lemma 15. *If $a \neq O$ or I is a central (i.e., complemented and neutral) element in L (which need not be complemented) and $r(x)$ is a normalized norm in L, then there is another normalized norm in L which differs from $r(x)$ either on a or on $\bar{a}$.*

For choose the numbers m and $\bar{m}$ so that $m+\bar{m}=1$, $m>0, \bar{m}>0$, and $m \neq \bar{m}$. Each $x=ax+\bar{a}x$. Define

$$\bar{R}(x) = mr(ax) + \bar{m}r(\bar{a}x).$$

Then $\bar{R}(O)=O$ and

$$\bar{R}(I) = mr(a) + \bar{m}r(\bar{a}) < r(a) + r(\bar{a}) = 1.$$

If $x<y$, then $ax \leqq ay$ and $\bar{a}x \leqq \bar{a}y$, and at least one of the equality signs does not hold. Thus

$$\bar{R}(x) = mr(ax) + \bar{m}r(\bar{a}x) < mr(ay) + \bar{m}r(\bar{a}y) = \bar{R}(y).$$

Also

$$\begin{aligned} \bar{R}(x) + \bar{R}(y) &= mr(ax) + \bar{m}r(\bar{a}x) + mr(ay) + \bar{m}r(\bar{a}y) \\ &= m[r(ax + ay) + r(axy)] + \bar{m}[r(\bar{a}x + \bar{a}y) + r(\bar{a}xy)] \\ &= m[r(a(x + y)) + r(axy)] + \bar{m}[r(\bar{a}(x + y)) + r(\bar{a}xy)] \\ &= \bar{R}(x + y) + \bar{R}(xy). \end{aligned}$$

Hence $\bar{r}(x)=\bar{R}(x)/\bar{R}(I)$ is a normalized norm over L. Now $\bar{R}(a)=mr(a)$ and $\bar{R}(\bar{a})=\bar{m}r(\bar{a})$; therefore $r(a)=\bar{r}(a)$ if and only if $m=\bar{R}(I)$, and $r(\bar{a})=\bar{r}(\bar{a})$ if and only if $\bar{m}=\bar{R}(I)$. Since $m \neq \bar{m}$, both cannot be true at the same time.

Corollary. *If the normalized quasi-norm $r(x)$ in the Boolean algebra L determines the σ-ideal $\mathfrak{a}$, then $r(x)$ is uniquely determined by $\mathfrak{a}$ if and only if $\mathfrak{a}$ is prime.*

For $L/\mathfrak{a}$ is a Boolean algebra, so that if it contains other elements beside O and I, they will be central elements; in such a case $r(x^*)$ and hence $r(x)$ would not be unique. Hence, if $r(x)$ is unique, $L/\mathfrak{a}=\{O^*, I^*\}$, and $\mathfrak{a}$ must be prime, since $L=\mathfrak{a}\cup\mathfrak{a}'$. Conversely, if $\mathfrak{a}$ is prime, $r(x^*)$ must be unique (note that $r(O^*)=0$ and $r(I^*)=1$) and $r(x)$ is uniquely determined by $\mathfrak{a}$.

Wilcox and Smiley(2) have given continuity conditions on a norm of L sufficient to assure isomorphism between the complete envelope $\bar{L}$ and the metrically complete envelope of L. Their conditions imply that if $a_\alpha \rightarrow A$ and $b_\alpha \rightarrow B$, then $r(a_\alpha) \rightarrow r(A)$, $a_\alpha + b_\alpha \rightarrow A+B$, and so on. With these conditions, MacNeille's problem(3) is answered in the positive: $\bar{L}$ is modular, since it possesses a norm. Because of the existence of O and I in our lattices, the W.S. conditions reduce to the following definition.

(2) Annals of Mathematics, (2), vol. 40 (1939), p. 309.

(3) See Birkhoff's *Lattice Theory*, American Mathematical Society Colloquium Publications, vol. 25, 1940, p. 146 (3).

A norm or quasi-norm in a lattice is called W.S. if for every $O \subset A \subseteq L$ we have

$$\text{l.u.b. } r(a) \text{ for } a \in A = \text{g.l.b. } r(b) \text{ for } b \in c_\tau A,$$

and

$$\text{g.l.b. } r(a) \text{ for } a \in A = \text{l.u.b. } r(b) \text{ for } b \in c_\sigma A.$$

THEOREM 11. *If the complemented modular L has a W.S. norm, then L is ideally irreducible if and only if its complete envelope, $\bar{L}$, is irreducible.*

For suppose that $\bar{L}$ is reducible with central element A and complement $\bar{A} \neq O, I$. Then $L \cap c_\sigma A$ and $L \cap c_\tau \bar{A}$ are normal. Since L is dense in $\bar{L}$,

$$A = \sum (L \cap c_\sigma A) = \prod (L \cap c_\tau A),$$

so that

$$L \cap c_\sigma A \neq O^\sigma, I^\sigma.$$

By Lemma 6 (which does not require complementation), $L \cap c_\sigma A$ and $L \cap c_\tau \bar{A}$ are complementary C-neutral ideals. Hence L is ideally reducible.

Conversely, if L is ideally reducible with $\mathfrak{a} \neq O^\sigma$, I^σ a normal C-neutral σ-ideal, then $\mathfrak{a}' = c_\sigma \bar{\mathfrak{a}}$ and $\bar{\mathfrak{a}}' = c_\tau \mathfrak{a}$ by Theorem 6 and its dual. Hence $\sum \mathfrak{a}' = \sum c_\sigma \bar{\mathfrak{a}} = \prod \bar{\mathfrak{a}}$ and $\prod \bar{\mathfrak{a}}' = \prod c_\tau \mathfrak{a} = \sum \mathfrak{a}$ by the following continuity argument and its dual: every $a' \leqq$ every $\bar{a}$, therefore $\sum \mathfrak{a}' = \sum c_\sigma \bar{\mathfrak{a}} \leqq \prod \bar{\mathfrak{a}}$; but

$$r(\sum \mathfrak{a}') = \text{l.u.b. } r(a') = \text{g.l.b. } r(\bar{a}) = r(\prod \bar{\mathfrak{a}});$$

hence $\sum \mathfrak{a}' = \prod \bar{\mathfrak{a}} = \bar{A}$, say, and $\sum \mathfrak{a} = \prod \bar{\mathfrak{a}}' = A$. Now if $a_\alpha' \uparrow A$ and $a_\alpha \uparrow A$, where $a_\alpha' \in \mathfrak{a}'$ and $a_\alpha \in \mathfrak{a}$, then $O = a_\alpha a_\alpha' \to A\bar{A}$. Consequently $A\bar{A} = O$ and dually $A + \bar{A} = I$.

By Lemma 12, $x(a_\alpha + a_\alpha') = xa_\alpha + xa_\alpha'$; hence $x(A + \bar{A}) = xA + x\bar{A}$ for every $x \in L$; therefore, finally, $X(A + \bar{A}) = XA + X\bar{A}$ for every $X \in \bar{L}$. Thus

$$X + Y = XA + X\bar{A} + YA + Y\bar{A},$$

and, since $\bar{L}$ is modular,

$$A(X + Y) = AX + AY,$$

and

$$\bar{A}(X + Y) = \bar{A}X + \bar{A}Y.$$

Consequently A and $\bar{A}$ are complementary central elements. $\bar{L}$ is therefore reducible, q.e.d.

THEOREM 12. *If L is a complemented modular lattice with a W.S. norm, $r(x)$, and $\mathfrak{a}$ is a normal C-neutral σ-ideal $\neq O^\sigma$ or I^σ, then there is another norm, $s(x)$, differing from $r(x)$ either on $\mathfrak{a}$ or on $\mathfrak{a}'$.*

For by Theorem 11, $\sum\mathfrak{a}$ is central in $\bar{L}$. Hence Lemma 15 applies in $\bar{L}$, and $s(x)$ specializes to another norm for L (which is a sublattice of $\bar{L}$).

COROLLARY. *If L is a complemented modular lattice with a W.S. norm, $r(x)$, then $r(x)$ is uniquely determined (except for linear transformation) if and only if L is ideally irreducible.*

For L fulfills von Neumann's continuity conditions (see Wilcox and Smiley). Hence, if L is irreducible, it is a continuous geometry, and von Neumann's theorem tells us that $r(x)$ is unique. The converse follows from Theorems 11 and 12.

THEOREM 13. *If the complemented modular L has the W.S. quasi-norm, $r(x)$ whose determined ideal is $\mathfrak{a}$, then $\mathfrak{a}$ is normal.*

For $\bar{\mathfrak{a}}' = c_\pi\mathfrak{a}$. Therefore

$$c_\sigma\bar{\mathfrak{a}}' = c_\sigma c_\pi\mathfrak{a} = \mathfrak{a}'',$$

and hence two applications of the continuity conditions yield

$$\text{l.u.b. } r(a'') = \text{g.l.b. } r(\bar{a}') = \text{l.u.b. } r(a) = 0.$$

Thus $r(a'')=0$, and $\mathfrak{a}''=\mathfrak{a}$.

THEOREM 14. *If L is a complemented modular lattice having $r(x)$ as a W.S. quasi-norm, and if the determined σ-ideal is $\mathfrak{a}$, then $\mathfrak{a}$ uniquely determines $r(x)$ if and only if it is maximal (prime for the π-ideal) among the normal C-neutral ideals.*

By the preceding results we have only to show that $r(a^*)$ is a W.S. norm in $L/\mathfrak{a}$. If $A^*\subseteq L/\mathfrak{a}$ and

$$x \in c_\pi \bigcup_{a^*\in A^*} a^*,$$

then $x\geqq$ every $a\in a^*\in A^*$, so that $x\in x^*\subseteq \bigcup c_\pi A^*$. Thus $c_\pi\bigcup A^*\subseteq\bigcup c_\pi A^*$. Similarly, $c_\sigma\bigcup A^*\subseteq\bigcup c_\sigma A^*$. Consequently,

$$\begin{aligned} \text{l.u.b. } r(a^*) \text{ for } a^* \in A^* &= \text{l.u.b. } r(a) \text{ for } a \in \textstyle\bigcup A^* \\ &= \text{g.l.b. } r(b) \text{ for } b \in c_\pi \textstyle\bigcup A^* \\ &\geqq \text{g.l.b. } r(b) \text{ for } b \in \textstyle\bigcup c_\pi A^* \\ &= \text{g.l.b. } r(b) \text{ for } b^* \in c_\pi A^*. \end{aligned}$$

The opposite inequality is obvious. Since the dual follows similarly, we have our result.

COLUMBIA UNIVERSITY,
NEW YORK, N.Y.

COMPLETE SETS OF LOGICAL FUNCTIONS

BY

WILLIAM WERNICK

A. In the two-valued calculus of sentences we denote variables by "p," "q," and their truth values by "1," "0." A truth function of two variables will be a function which assumes the value 1, or 0, depending on the truth values of the variables p, q, in its argument. Since each variable may assume independently the value 1, 0, there are 2^2 possible arguments for each function. A particular function is completely determined if we assign to each of its 4 possible arguments a value which may be independently 1, or 0, so there are altogether 2^{2^2}, i.e., 16 possible functions in this calculus. Some of these functions are familiar to us by name: conjunction, implication, negation, etc. When we wish to speak of all 16 functions collectively, we shall call the totality "C."

Sheffer has shown (1) that we may take one of these as undefined and define formally in terms of this one all the other functions of the calculus. He also pointed out that this function, which he called "stroke," could be given a dual interpretation, that is, there are two functions in C, mutually dual, which, alone, can define all the other 15 functions in C. Whitehead and Russell use the pair of undefined functions *negation* and *disjunction* to generate this calculus; Hilbert and Ackerman show(2) that this pair will suffice, as well as the pairs *negation* and *conjunction*, or *negation* and *implication*. They show also that the pair *negation* and *equivalence* is insufficient to generate the calculus, being incapable, in particular, of defining *conjunction*.

B. If we take as undefined the set of functions $\{\alpha, \beta, \cdots, \delta\}$ we shall look for such sets which have the following properties:

1. All the 16 functions of C may be formally(3) defined in terms of functions of this set. (Sufficiency.)

2. No function of this set may be defined in terms of some or all the remaining functions of the set. (Non-redundancy.)

If a set of undefined functions is both sufficient and non-redundant, it will be called "complete." Thus the set *negation* and *disjunction* is a complete

Presented to the Society, February 24, 1940; received by the editors March 11, 1940 and, in revised form, December 10, 1940. This paper is the substance of the author's doctoral thesis at New York University, February, 1941. The author wishes to extend thanks to Dr. J. C. C. McKinsey for his interest in this investigation.

(1) H. M. Sheffer, these Transactions, vol. 14 (1913), pp. 481–488.

(2) Hilbert and Ackerman, *Grundzüge der theoretischen Logik*, 2d edition, 1938, p. 8 ff.

(3) A definition will be called "formal" if the definiens is simply an expression built up by the use of sentential connectives (our functions) and sentential variables "p," "q." All other definitions are non-formal.

set. This paper will enumerate all complete sets and show that they may contain one, two, or three functions, but not more. Post points out(4) the possibility of generalizing the Principia development by using other primitive (i.e., generating) functions than *negation* and *disjunction*, but does not discuss complete sets of such functions(5).

C. We may start by listing the 16 functions thus:

TABLE I

p	q	a	b	c	d	e	f	g	h	a'	b'	c'	d'	e'	f'	g'	h'
1	1	0	1	0	0	0	1	1	1	1	0	1	1	1	0	0	0
1	0	0	0	1	0	0	1	0	0	1	1	0	1	1	0	1	1
0	1	0	0	0	1	0	0	1	0	1	1	1	0	1	1	0	1
0	0	0	0	0	0	1	0	0	1	1	1	1	1	0	1	1	0

We shall take our arguments in what follows so that the truth values of p, q, will always be in the order 1, 1; 1, 0; 0, 1; 0, 0. The names given to these functions are immaterial, and the use of "a'," "b'," etc., instead of "i,", "j," etc., has only mnemonic significance. We may note some remarks on this table:

1. The function $b(p, q)$, defined above, may be recognized as the truth function "logical product," $p \cdot q$, which is true only when p and q are both true. Interpretations of all the 16 functions, in the more familiar notation of symbolic logic, are given here, with the emphasis on familiarity, and not on economy of symbols:

TABLE II

a: $p \cdot \sim p$	e: $\sim p \cdot \sim q$	a': $p \vee \sim p$	e': $p \vee q$
b: $p \cdot q$	f: p	b': $\sim(p \cdot q)$	f': $\sim p$
c: $\sim(p \supset q)$	g: q	c': $p \supset q$	g': $\sim q$
d: $\sim(q \supset p)$	h: $p \equiv q$	d': $q \supset p$	h': $\sim(p \equiv q)$

2. The function $f(p, q)$ need never be an undefined function, since it is immediately definable with no additional machinery. The function $f(p, q)$ is exactly p, being true when and only when p is true, as may be seen from Table I. Likewise $g(p, q) = q$. This means in any truth function involving p or q, we may replace p by $f(p, q)$, or q by $g(p, q)$(6).

3. We introduce here the notion of symmetric functions. The function α

(4) E. L. Post, *Introduction to a general theory of elementary propositions*, American Journal of Mathematics, vol. 43 (1921), pp. 163–185.

(5) His concept of a complete system should not be confused with our concept of a complete set of functions. He calls a system of truth tables "complete" if it contains all possible truth tables of the logic. Thus, the listing of all 16 functions of C in Table I below would form a complete system in Post's notation. He uses the notion of redundancy of a set of generating functions only implicitly (when he speaks of the "order" of a system).

(6) We use "or" throughout this paper to mean either alternative or both.

will be called symmetric to the function β, if the following four equations all hold:

$$\alpha(1, 1) = \beta(1, 1), \qquad \alpha(0, 1) = \beta(1, 0),$$
$$\alpha(1, 0) = \beta(0, 1), \qquad \alpha(0, 0) = \beta(0, 0).$$

Note that if $\alpha(1, 0) = \alpha(0, 1)$, then α is self-symmetric, i.e., the value is independent of the order of the arguments. If, in Table II, we replace "p" by "q," and "q" by "p" in the definition of any particular function, we obtain the definition of its symmetric function. If there resulted no change (due to the commutative nature of the connectives $\cdot$, $\vee$, and $\equiv$), then that function was self-symmetric. The reader can easily verify the list of self-symmetric functions: a, b, e, h, a', b', e', h'; and the others symmetric in pairs: cSd; fSg; $c'Sd'$; and $f'Sg'$.

4. Since we are dealing only with functions of two arguments, any function of p alone will be considered as a function of both p and q which is independent of q; e.g., $f'(p, q)$ is the function "p is false," and is independent of q. Similarly for a function of q alone.

5. The fact that our functions are defined for arguments of the form p, q, as p, q go through their ordered changes, means that it is unnecessary to consider functions of q, p. For example, $c'(p, q)$ is our definition for "p implies q," and we may not use $c'(q, p)$ to mean "q implies p," but must use some function of p, q, since our definitions are set up on the agreement that the first place in the table is the value of the function for $p=1$ and $q=1$, the second place is the value for $p=1$ and $q=0$, etc. Introduction of arguments of the form q, p would disrupt our definitions. They are, moreover, superfluous, since any function of q, p has an equivalent function of p, q.

In view of the theorem proved in §H1 we remark, that so far as completeness is concerned, the functions d, d', g, g' are equivalent to c, c', f, f', respectively. We shall omit, from now on, further unnecessary reference to d, d', g, g' when they are to be undefined functions. They continue to play their roles as functions of C to be defined by others.

D. We introduce now an abbreviated notation which will greatly reduce the amount and complexity of the material to follow. Consider the function of functions $e[h(p, q), e'(p, q)]$. As p, q go through their ordered changes $(1, 1; 1, 0; 0, 1; 0, 0)$ we get the values 0, 0, 0, 0.

But this set of values is the same set we would have obtained if we had let p, q go through their ordered changes for the function $a(p, q)$ (see Table I). This establishes the identity $e[h(p, q), e'(p, q)] = a(p, q)$ for all values of p, q. This identity will now be written simply as $e(h, e') = a$.

E. We consider, from now on, only functions of the form $\alpha(\beta, \gamma)$, where α, β, γ are any of the 16 functions in C. The functions f, g play a special role here, as noted in §C, since $\alpha(f, g) = \alpha(p, q)$. While we do not define $\alpha(p, p)$, we have defined $\alpha(f, f)$, which is equivalent to it.

Consider the sets of functions defined thus:

$\{\alpha\}_0$: f, g;

$\{\alpha\}_1$: $\alpha(f, f)$; $\alpha(f, g)$; $\alpha(g, f)$; $\alpha(g, g)$;

. . . . ;

$\{\alpha\}_i$: all functions in $\{\{\alpha\}_{i-1}\}$ with f or g or both replaced by any functions in $\{\{\alpha\}_k\}$, $k < i - 1$.

Clearly, $\{\alpha\}_1$ will contain the function $\alpha(p, q)$, since this is equal to $\alpha(f, g)$, and may, from considerations in §C, contain other functions of p, q; e.g., if α is the function e then $f' = e(f, f)$. Also, $\{\alpha\}_2$ may contain functions not in $\{\alpha\}_1$; and $\{\alpha\}_3$ may contain functions not in $\{\alpha\}_2$ or in $\{\alpha\}_1$. (It is understood that all functions that occur in the above constructions can be reduced to, and are replaced by, the equivalent single functions of C.) This accretion of new functions must soon stop, since there are only 16 different functions altogether. The totality of all (reduced) functions of p, q, in all the sets $\{\{\alpha\}_i\}$, where the i is sufficiently large to insure that no new functions can be introduced by increasing i, will be called the field of the function α, indicated by "$F(\alpha)$;" thus, from the beginning of this paragraph we have: f' is in $F(e)$. The process of enumerating a field is always finite, and there are immediately evident to an investigator special theorems for each particular function that reduce the work of enumeration. Since we are concerned only with the totality of functions in $F(\alpha)$, in extension and without repetitions, we can immediately discard repetitions of functions in C which we have already found in $\{\{\alpha\}_{i-1}\}$ before investigating $\{\alpha\}_i$(7). The problem: "What functions of C are in $F(\alpha)$?" is seen to be equivalent to the problem: "What functions in the calculus of sentences can be defined, if we take a particular function α as undefined?"(8).

F. If there are two undefined functions α, β, we have analogously the sets of functions

$\{\alpha, \beta\}_0$: f, g;

$\{\alpha, \beta\}_1$: $\{\alpha\}_1 + \{\beta\}_1$;

. . . . ;

$\{\alpha, \beta\}_i$: all functions in $\{\{\alpha, \beta\}_{i-1}\}$ with f or g or both replaced by any function in $\{\{\alpha, \beta\}_k\}$, $k < i - 1$.

(7) E. Żyliński, *Some remarks concerning the theory of deduction*, Fundamenta Mathematicae, vol. 7 (1925), pp. 203–209 also discusses, with different terminology, the notion of a field. He asks, in effect, for the field for each function and indicates the results given in Table III below, but does not discuss fields of two or more functions.

(8) Since f, and g, can be defined without the aid of any function, they properly belong in every field. These two functions can be formally defined in terms of 14 of our 16 functions, and cannot be formally defined in terms of a, or a'; nevertheless, we place f and g in $F(a)$, and in $F(a')$, since we have, for example, the definitions: $a(f, g) = a = a(p, q)$; $f(f, g) = p = f(p, q)$; $g(f, g) = q = g(p, q)$; with only the function a as an undefined function.

Note that $\{\alpha, \beta\}_2$ certainly includes $\{\alpha\}_2+\{\beta\}_2$ but may easily include more functions. The totality of functions (reduced and without repetitions) in all the sets $\{\{\alpha, \beta\}_i\}$ will be called the field of the functions α, β, indicated by "$F(\alpha, \beta)$." As before, the process of enumerating a field is always finite and becomes quickly simplified in any particular case by the application of simple particularized theorems which apply to each pair. As before, the problem: "What functions are in $F(\alpha, \beta)$?" is equivalent to the problem: "What functions in the calculus of sentences can be defined in terms of only two given undefined functions?"

There are corresponding definitions for sets and fields of three undefined functions α, β, γ; of four undefined functions, etc. We may point out here that Post (loc. cit.) discusses the totality of functions formally definable in terms of a set of u arbitrary functions $\{f_1, \cdots, f_u\}$, where f_i is a function of m_i arguments. This totality of functions is called F, and corresponds to our notion of a field. He does not concern himself with redundancy among the $\{f_i\}$.

Since we are concerned with sufficient sets of functions, we shall try to find functions whose field is the entire set of 16, i.e., whose field is C. Also, since we are looking for non-redundant sets of functions, we shall discard a set $\{\alpha, \beta, \gamma, \delta, \cdots\}$ if, for example, α is in $F(\beta, \gamma, \delta, \cdots)$.

G. We now prove that the functions e and b' are the only single functions that form complete sets(9). Obviously, the set $\{e\}$ is non-redundant, as is the set $\{b'\}$. If we are seeking a single function α, so that $F(\alpha)=C$, it must satisfy the condition that $\alpha(1, 1)=0$, and $\alpha(0, 0)=1$(10). For, if $\alpha(1, 1)=1$ then any function of α would have as its first pair of arguments the truth values 1, 1, and hence, for those arguments, would have the *value* 1, whereas half the functions in C have the value 0 for the first pair of arguments. Similar reasoning assures that $\alpha(0, 0)$ must equal 1. Looking through Table I for functions of the form 0, $-$, $-$, 1; we find only the functions e, b', f', g'. A function of one variable cannot define a function really dependent on two variables, so that our function α which is to define all the functions in C may not be independent of one of its arguments. But we quickly establish the fact that f' is independent of q, and g' is independent of p (since $f'(1, 1)=f'(1, 0)$, and $f'(0, 1)=f'(0, 0)$; and similarly for g'). Hence e and b' are the only possible functions.

Since Sheffer (loc. cit.) has already shown that these two functions are sufficient, this completes the proof that the only complete sets consisting of single functions are $\{e\}$, and $\{b'\}$.

H. If we look for complete sets consisting of two functions, the realm of

(9) These two functions are Sheffer's stroke function and its dual. He states (loc. cit.), in effect, that these two sets are complete. Żyliński (loc. cit.) states, in effect, but does not prove, that these are the only complete sets of single functions. The proof is given here in full because it is typical.

(10) This is essentially the property of negation.

investigation broadens considerably. There seem to be $C_{16,2}$, i.e., 120 possible pairs of functions to investigate, but some obvious considerations will quickly reduce this number. Any pair containing the functions e, or b' is, by §G, immediately redundant, since all functions are in $F(e)$, and in $F(b')$. Also, any pair containing f or g is redundant, since, from the definition of a field, these functions are in every field. Not all of the remaining $C_{12,2}$, i.e., 66 pairs of functions are essentially different, since, in view of the theorem on symmetric functions mentioned in §C above, any function pair containing d, d', or g' is equivalent, as far as definability is concerned, to the same pair with d, d', or g' replaced by c, c', or f', respectively. This leaves only 36 pairs of functions to investigate. To find the redundant functions, we determine first the fields of single functions as per Table III(11).

TABLE III

$F(a)=a, f, g$	$F(a')=a', f, g$
$F(b)=b, f, g$	$F(b')=C$
$F(c)=a, b, c, d, f, g$	$F(c')=f, g, a', c', d', e'$
$F(d)=a, b, c, d, f, g$	$F(d')=f, g, a', c', d', e'$
$F(e)=C$	$F(e')=f, g, e'$
$F(f)=f, g$	$F(f')=f, g, f', g'$
$F(g)=f, g$	$F(g')=f, g, f', g'$
$F(h)=f, g, h, a'$	$F(h')=a, f, g, h'$

We note some remarks on Table III:

1. *Symmetry*. The identity of the fields for certain functions was predicted by a theorem mentioned earlier: "If α is symmetric to β, then $F(\alpha)=F(\beta)$," which we now prove.

If α is symmetric to β, and ϵ is in $F(\alpha)$, then there must exist functions γ, δ, in $F(\alpha)$ so that $\epsilon=\alpha(\gamma, \delta)$. From the definition of symmetry, we must also have $\epsilon=\beta(\delta, \gamma)$, and ϵ, therefore, will be in $F(\beta)$ if γ, δ are in $F(\beta)$. Applying to γ, δ the same argument we have just applied to ϵ, we form a recursion in a very few steps to $\{\alpha\}_0$, and $\{\beta\}_0$. But these are, by definition, identical, consisting of the functions f, and g; hence the theorem.

2. *Transitivity*. The reader can verify the rather obvious theorem: "If α is in $F(\beta)$, and β is in $F(\gamma)$, then α is in $F(\gamma)$." In terms of definability, this theorem states that if β can define α, and γ can define β, then γ can define α. As a corollary to this theorem on transitivity, which seems intuitively evident, we can derive the theorem given above under symmetry; for, if α is symmetric to β then clearly α is in $F(\beta)$.

3. *Duality*. The function α is called the dual of the function β, notation $\alpha=\bar{\beta}$, if, for every p and q, we have $\alpha(p, q)=\sim\beta(\sim p, \sim q)$. We use "$\sim$" in

(11) For purposes of completeness, we indicate the fields for each of the 16 functions in C, though we shall not need all these fields for the remainder of this paper. The results in this table also appear in Żyliński's paper (loc. cit.).

its usual sense, so that $\sim 1=0$, and $\sim 0=1$. We state the useful and simple theorem: "If $\alpha=\bar{\beta}$, then $\beta=\bar{\alpha}$," whose proof is left to the reader.

We list here the relations of duality that exist among the functions in C:

TABLE IV

$a=\overline{a'}$	$c=\overline{d'}$	$e=\overline{b'}$	$g=\bar{g}$	$f'=\overline{f'}$
$b=\overline{e'}$	$d=\overline{c'}$	$f=\bar{f}$	$h=\overline{h'}$	$g'=\overline{g'}$

This table may be verified in the following way, suggested by Żyliński (loc. cit.). In Table I, replace "0" by "1," and "1" by "0" *everywhere*; then the matrix definition under each function designation will be that of the dual of that function. This amounts to interchanging "0" and "1" in the *definitions*, and then writing the columns upside down. The equation we have given to define duality is, of course, equivalent to this transformation.

From Tables III and IV we could easily verify the theorem: "If α is in $F(\beta)$, then $\bar{\alpha}$ is in $F(\bar{\beta})$," the usefulness of which will be more apparent if we state it in the form: "If $\beta=\bar{\alpha}$, and $F(\alpha)=\gamma_1, \cdots, \gamma_n$, then $F(\beta)\overline{\gamma_1}=, \cdots, \overline{\gamma_n}$." To prove it directly, we first prove that if $\alpha=\beta(\gamma, \delta)$ then $\bar{\alpha}=\bar{\beta}(\bar{\gamma}, \bar{\delta})$. Our hypothesis, written out in full is: "For every p, and q, we have the identity $\alpha(p, q)=\beta(\gamma(p, q), \delta(p, q))$." Replace "$p$" by "$\sim p$," "$q$" by "$\sim q$," and preface both sides of the equation by "$\sim$." This gives us

$$(1) \qquad \sim \alpha(\sim p, \sim q) = \sim \beta(\gamma(\sim p, \sim q), \delta(\sim p, \sim q)).$$

From the definition of dual functions, we have $\bar{\alpha}(p, q)=\sim\alpha(\sim p, \sim q)$, or, by prefacing both sides of this equation by "$\sim$," and dropping "$\sim\sim$," $\sim\bar{\alpha}(p, q)=\alpha(\sim p, \sim q)$. Applying these equations to (1), we have

$$(2) \qquad \bar{\alpha}(p, q) = \sim \beta(\sim \bar{\gamma}(p, q), \sim \bar{\delta}(p, q)),$$

which can be abbreviated to

$$(3) \qquad \bar{\alpha} = \sim \beta(\sim \bar{\gamma}, \sim \bar{\delta}).$$

But, again, from the definition of duality, we have $\bar{\beta}(p, q)=\sim\beta(\sim p, \sim q)$, and by replacing "$p$" by "$\bar{\gamma}$," and "$q$" by "$\bar{\delta}$," we have $\bar{\beta}(\bar{\gamma}, \bar{\delta})=\sim\beta(\sim\bar{\gamma}, \sim\bar{\delta})$; therefore, finally $\bar{\alpha}=\bar{\beta}(\bar{\gamma}, \bar{\delta})$.

To say that α is in $F(\beta)$ is to say that there is an equation $\alpha=\beta(\gamma, \delta)$, where γ, δ are in $F(\beta)$. But if there is such an equation, we have just shown that there is an equally valid equation $\bar{\alpha}=\bar{\beta}(\bar{\gamma}, \bar{\delta})$; that is, $\bar{\alpha}$ is in $F(\bar{\beta})$, provided only that $\bar{\gamma}$, $\bar{\delta}$ are in $F(\bar{\beta})$. We have, now, a simple recursion, applying to γ, δ the same arguments as to α, and leading to $\{\beta\}_0$ and $\{\bar{\beta}\}_0$, that is, our theorem depends on: "If the functions in $\{\beta\}_0$ are in $F\{\beta\}$, then the functions in $\{\bar{\beta}\}_0$ are in $F(\bar{\beta})$;" which is, of course, a true theorem, since $\{\epsilon\}_0=f, g$, for any ϵ. Hence our main theorem.

These properties of duality may be used to check our results later on, since

all definitions may be dualized and yield equally valid definitions. (If, e.g., $\{\alpha, \beta\}$ is a complete set, so must $\{\bar{\alpha}, \bar{\beta}\}$ be.)

4. *Identity of fields.* Since $F(c)=F(d)$; $F(e)=F(b')$; $F(f)=F(g)$; $F(c')=F(d')$; $F(f')=F(g')$; there are just 11 different fields generated by a single function. Post (loc. cit.) has investigated systems of functions definable by functions of more than two variables. He states, in effect, that there are just 66 different fields which can be generated by functions of fewer than four arguments, and there are 8 infinite families which can be generated by functions of four or more arguments. He does not discuss, however, such concepts as duality, symmetry, and transitivity.

We shall not verify the enumeration of all fields in Table III, but shall indicate the method by which this may be done.

There are two steps in the verification of $F(\alpha)$. First we must prove that the functions listed as belonging to $F(\alpha)$ are actually in $F(\alpha)$, which could be done immediately by giving the definitions in terms of α, or of functions defined in terms of α, etc. (using transitivity). Second, we must show that no other function belongs in that field, i.e., that the field is closed. We may use special theorems for each function, but we can always use this straightforward method of proving closure: If δ is in $F(\alpha)$, then there must be a definition $\delta=\alpha(\beta, \gamma)$, where β, γ are in $F(\alpha)$. If we have already shown that there are, say, six functions in $F(\alpha)$, we can substitute for β, or γ, any of these six functions, giving, in this case, 36 possible pairs of argument functions. Closure is proved if the values resulting from these 36 arguments are among the six functions already shown to be in $F(\alpha)$. It is never necessary to investigate the result of every substitution, since special theorems give the results for many classes of substitutions.

The verification of most of the fields is straightforward, but we verify $F(c)$ in detail to illustrate another typical method, $F(c)=a, b, c, d, f, g$; $a=c(f, f)$; $b=c(f, c)$; $d=c(g, f)$. If α is any of the above six functions in $F(c)$, then $\alpha(0, 0)=0$ (see Table I). Therefore (see §G) any function of these six functions in $F(c)$ will reduce, for $p=0$, $q=0$, to the value 0. Hence $F(c)$ consists of at most the eight functions in C for which the argument 0, 0 gives the value 0. To prove closure, we must now show that e' and h', the only other functions which have this property, are not in $F(c)$. From the definition of c, $c(0, x)=0$, for all x, and $c(1, y)=1$ only when $y=0$. Therefore, $c(\alpha, \beta)$ can have the value 1 at most as many times as α has the value 1, and then at the same places. If e' is in $F(c)$ then there must exist functions α, β, in $F(c)$ such that $e'=c(\alpha, \beta)$, where, furthermore, α is of the form 1, 1, 1, x; α, in turn, requires in its definition another function with the same form, etc. Therefore, there must be in $\{c\}_0$ a function of the form 1, 1, 1, x; which is not the case. Therefore, e' is not in $F(c)$. Similarly, h' in $F(c)$ requires eventually that $\{c\}_0$ contain a function of the form x, 1, 1, y; which it does not contain. Therefore, h' is not in $F(c)$. This concludes the proof that $F(c)$ is closed as given.

From Table III we may now easily select the redundant pairs of functions. We indicate them, 67 in all, in the following table([12]):

TABLE V

a, c	b, c	h, a'	c', d'	f', g'
a, d	b, d	a', c'	c', e'	
a, h'	c, d	a', d'	d', e'	
and, all the 54 possible pairs of functions with e, or b', or f, or g; as one of the pair				

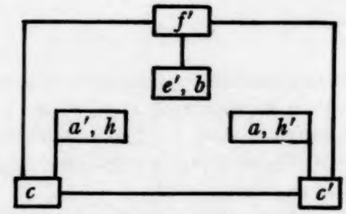
The 54 pairs of redundant functions are not written down, since they are obvious, but they, and the 13 pairs directly above, comprise *all* the redundant function pairs, i.e., pairs of which one function (not either one) may be taken as undefined, and the other will be immediately definable. Of these first 13 pairs, only 6 are essentially different and must be subtracted from the 36 pairs mentioned at the beginning of §H (e.g., the pairs a, c; a, d are not essentially different, since cSd, and $F(a, c) = F(a, d)$).

I. We now have left to consider, from the total of 120 pairs of functions indicated in §H, only 30 essentially different pairs which are not redundant, but whose sufficiency is still undecided. We determine, by the method shown in §F, the fields for each of these 30 pairs, and find that there are exactly 9 pairs which form sufficient sets. Since these are selected from non-redundant pairs, these 9 sets are, in fact, complete sets of two functions, and are listed below. This listing answers completely the question: What pairs of functions may be used to define the calculus of sentences?

TABLE VI

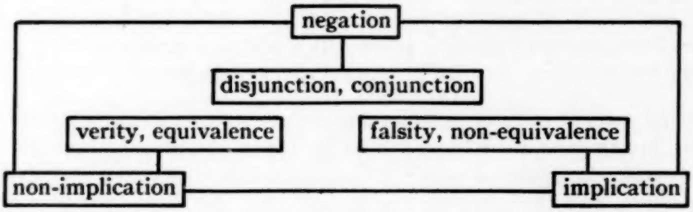
a, c'	c, a'	c', f'
b, f'	c, c'	c', h'
c, h	c, f'	e', f'

This table may be represented by the following diagram:



([12]) We shall leave out, in the text and tables from now on the braces around functions even though we mean them to be generating sets, whether complete or not.

Any function in one box may be paired with a function in another box to which it is joined by a line, to form a complete two-function set. The preceding diagram can be translated in an obvious way (see Table II) to yield the following diagram, of some interest in the calculus of sentences, since it gives all essentially different pairs of functions which can define this calculus:



It is possible to show the sufficiency of these pairs without actually enumerating the fields for each pair, by applying the theorem stated in the remark on transitivity, following Table III. Since we have shown that

$$F(e) = F(b') = C,$$

as soon as we can show that e or b' is in $F(\alpha, \beta)$, it follows immediately that

$$F(\alpha, \beta) = C.$$

We do this now for all the pairs in Table VI. In most cases, a single definition will suffice, but we need at most two defined functions. The definitions given here are not the only possible definitions for the function actually defined, and we define e or b', depending on which comes first to hand, or seems easier.

TABLE VII

a, c': $f'=c'(f, a)$, $b'=c'(c', f')$	c', f': $b'=c'(g, f')$
c, a': $f'=c(a', f)$, $e=c(f', g)$	c', h': $b'=h'(f, c')$
b, f': $b'=f'(b, f)$	e', f': $e=f'(e', f)$
c, h: $b'=h(f, c)$	
c, c': $e=c(c', g)$	
c, f': $e=c(f', g)$	

If Tables VI and VII had included *all* complete pairs (i.e., those given, and those obtainable from them by replacing c, c', f' by d, d', g', respectively), we could have used a duality theorem to check our tables; e.g., from the complete set c, h and its definition of b', we could obtain the dual (complete) set d', h' and its definition of e.

J. There still remain, of the 30 non-redundant pairs mentioned in §I, a total of 21 pairs of functions which form non-redundant but insufficient sets. We list them in Table VIII below:

TABLE VIII

a, b	a, e'	b, a'	b, h'	h, c'	h, h'	a', h'
a, h	a, f'	b, c'	c, e'	h, e'	a', e'	e', h'
a, a'	b, h	b, e'	c, h'	h, f'	a', f'	f', h'

The assertion here that these pairs are insufficient can only be substantiated if we list and verify the fields for each of the 21 pairs in Table VIII, which we shall do in more detail than we did for single functions in Table III.

We shall list the functions in $F(\alpha, \beta)$ thus: $F(\alpha, \beta) = \gamma_1, \cdots, \gamma_p; \delta_1, \cdots, \delta_q; \epsilon_1, \cdots, \epsilon_r$, where γ_i is in $F(\alpha)$, δ_i is in $F(\alpha)+F(\beta)$ but not in $F(\alpha)$, and ϵ_i is in $F(\alpha, \beta)$ but not in $F(\alpha)+F(\beta)$. We shall give definitions only for the functions ϵ_i([13]), since definitions for γ_i and δ_i have already been indicated in §H. Proofs of closure to show that no more functions than those listed belong to $F(\alpha, \beta)$ are given after the listing of the fields in Table IX below:

TABLE IX

1. $F(a, b) = a, f, g; b$	
2. $F(a, h) = a, f, g; h, a'; f', g', h'$	$f' = h(f, a);\ g' = h(g, a);\ h' = h(h, a)$
3. $F(a, a') = a, f, g; a'$	
4. $F(a, e') = a, f, g; e'$	
5. $F(a, f') = a, f, g; f', g'; a'$	$a' = f'(a, f)$
6. $F(b, h) = b, f, g; h, a'; c', d', e'$	$c' = h(f, b);\ d' = h(g, b);\ e' = h(b, h)$
7. $F(b, a') = b, f, g; a'$	
8. $F(b, c') = b, f, g; a', c', d', e'; h$	$h = b(c', d')$
9. $F(b, e') = b, f, g; e'$	
10. $F(b, h') = b, f, g; a, h'; c, d, e'$	$c = h'(f, b);\ d = h'(g, b);\ e' = h'(f, d)$
11. $F(c, e') = a, b, c, d, f, g; e'; h'$	$h' = c(e', b)$
12. $F(c, h') = a, b, c, d, f, g; h'; e'$	$e' = h'(g, c)$
13. $F(h, c') = f, g, h, a'; c', d', e'; b$	$b = h(g, d')$
14. $F(h, e') = f, g, h, a'; e'; b, c', d'$	$b = h(h, e');\ c' = h(g, e');\ d' = h(g, b)$
15. $F(h, f') = f, g, h, a'; f', g'; a, h'$	$a = f'(a', f);\ h' = f'(h, f)$
16. $F(h, h') = f, g, h, a'; a, h'; f', g'$	$f' = h'(g, h);\ g' = h'(f, h)$
17. $F(a', e') = f, g, a'; e'$	
18. $F(a', f') = f, g, a'; f', g'; a$	$a = f'(a', f)$
19. $F(a', h') = f, g, a'; a, h'; h, f', g'$	$h = h'(a', h');\ f' = h'(a', f);\ g' = h'(a', g)$
20. $F(e', h') = f, g, e'; a, h'; b, c, d$	$b = h'(e', h');\ c = h'(f, b);\ d = h'(g, b)$
21. $F(f', h') = f, g, f', g'; a, h'; h, a'$	$h = h'(f', g);\ a' = f'(a, f)$

We may prove closure for several of these fields at once. Notice that many of the fields are identical and often consist of eight functions. Consider these:

([13]) We list the complete fields and definitions for *all* functions ϵ_i, including even the defined functions d, d', g', which, we know, must appear in any field in which their respective symmetrics c, c', f' appear.

$$F_1: \quad a, b, c, d, f, g, e', h',$$
$$F_2: \quad b, f, g, h, a', c', d', e',$$
$$F_3: \quad a, f, g, h, a', f', g', h'.$$

If we examine the definitions of these functions in Table I, we may note:

(i) F_1 consists of all functions $\{\alpha\}$ in C for which $\alpha(0, 0)=0$;

(ii) F_2 consists of all functions $\{\alpha\}$ in C for which $\alpha(1, 1)=1$;

(iii) F_3 consists of all functions $\{\alpha\}$ in C for which the sum of the four constants defining α is even, i.e., F_3 consists of all even functions.

We shall prove a general theorem: "If $\alpha, \beta, \cdots, \delta$, are all in F_i $(i=1, 2, 3)$, then every function in $F(\alpha, \beta, \cdots, \delta)$ is in F_i."

F_1: From the remark at the beginning of §G, we know that if $\alpha(0, 0)=0$, and if ϵ is in $F(\alpha)$, then we can never have $\epsilon(0, 0)=1$. That is, if α is in F_1, then every function in $F(\alpha)$ is also in F_1. This argument holds as well if there are two functions to generate the field: if α and β are in F_1, then any function defined in terms of α and β (i.e., any function in $F(\alpha, \beta)$) must always have for $p=0$, $q=0$, the value 0, i.e., must be in F_1. The extension to several functions, $\alpha, \beta, \cdots, \delta$, all of which are in F_1 is immediate.

F_2: Referring to the same remark as above, for functions α for which $\alpha(1, 1)=1$, we derive in the same way an analogous conclusion: if, for $p=1$, $q=1$, the functions $\alpha, \beta, \cdots, \delta$, take the value 1, then every function in $F(\alpha, \beta, \cdots, \delta)$ must also have this property. But the condition of this statement is precisely that $\alpha, \beta, \cdots, \delta$ all be in F_2, and the conclusion that every function in $F(\alpha, \beta, \cdots, \delta)$ is also in F_2.

F_3: Consider functions of the form $\alpha(\beta, \gamma)$, where α, β, γ may be replaced by any function in F_3. We must show that the result is always a function in F_3. Consider replacements of α by functions in F_3. We shall use known properties of these functions in the form of theorem-equations without proof (which would follow immediately from their definitions). We have for every β, γ, the following:

$$a(\beta, \gamma) = a, \qquad a'(\beta, \gamma) = a',$$
$$f(\beta, \gamma) = \beta, \qquad f'(\beta, \gamma) = \beta',$$
$$g(\beta, \gamma) = \gamma, \qquad g'(\beta, \gamma) = \gamma'.$$

If β, γ are in F_3, clearly β' and γ' are also in F_3, and the results of all the above replacements are also in F_3; so we need now investigate the cases $h(\beta, \gamma)$, and $h'(\beta, \gamma)$. But again, from the definitions we have

$$h'(\beta, \gamma) = (h(\beta, \gamma))',$$

so we need now investigate only $h(\beta, \gamma)$. Note that this involves more than $F(h)$, since β or γ need not be in $F(h)$, but must be in F_3, which contains more than $F(h)$.

We have also the theorem-equation

$$h(\beta, \gamma') = (h(\beta, \gamma))',$$

which makes it unnecessary to investigate primed arguments; so we need to investigate only replacements of β, γ by a, f, g, h. And now, finally, these last theorems:

$$h(\gamma, a) = h(a, \gamma) = \gamma'; \qquad h(\beta, \beta) = a';$$

make it necessary to test only the three possibilities $h(f, g)$, $h(f, h)$, $h(g, h)$. But all these are, of course, in $F(h)$ which is contained in F_3, their primes are likewise in F_3, and the last part of our general theorem is thus established.

This last part may be stated: "An odd function can never be in a field generated only by even functions," and can be used for checking many fields without proving closure. For example: b is not in $F(h, f')$, since h, f' are even and can only generate an even field, whereas b is odd([14]).

We may obtain some immediate corollaries:

COROLLARY 1. *Every complete set must contain at least one odd function.*

COROLLARY 2. *A necessary condition that the set* $\alpha, \beta, \cdots, \delta$ *be complete is that not all the functions* $\alpha, \beta, \cdots, \delta$ *be in some* F_i.

COROLLARY 3. *Consider the sets of functions:*

$$F_4\colon\ a, f, g, a', f', g'; \qquad F_5\colon\ a, a';$$

which are subsets of F_3. *We have, from the last part of the proof above, the corollary that the general theorem includes the cases* $i=4, 5$.

We derive now some important consequences of this general theorem. Among the 21 pairs in Table VIII we notice many pairs which have both their members in some F_i, and hence come within the scope of this theorem. In particular, the four pairs b, h'; c, e'; c, h'; e', h' are wholly in F_1 and hence, by the theorem, their fields are at most F_1. But we have already shown that their fields are at least F_1; therefore, their fields are exactly F_1. Similarly, there are four pairs which are wholly in F_2: b, h; b, c'; h, c'; h, e'; to which we may apply exactly the same method. We have shown that their fields are at least F_2, and the theorem states that their fields are at most F_2; therefore, their fields are exactly F_2. The theorem is still applicable to the five function pairs wholly in F_3; a, h; h, f'; h, h'; a', h'; f', h'; whose fields are therefore proved to be closed.

We have, with the general theorem, proved closure for 13 of our 21 fields. Corollary 3 may be used to prove closure for three more fields. In particular,

---

(14) That is, the pair of functions *negation* and *equivalence* are incapable of defining *conjunction*, as stated by Hilbert and Ackerman (loc. cit.).

the two pairs a, f' and a', f' are in F_4; therefore, by the usual argument, their fields are exactly F_4. Also, the pair a, a' is in F_5; therefore, its field is exactly F_5.

This still leaves the five pairs a, b; a', b; a, e'; a', e'; b, e'; for whose fields we must prove closure. We know, for any α, β, that

$$a(\alpha, \beta) = a; \qquad a'(\alpha, \beta) = a'.$$

Also, from the definitions of b, and e', which are self-symmetric, that $b(a, \alpha) = a$, $b(a', \alpha) = \alpha$, $e'(a, \alpha) = \alpha$, and $e'(a', \alpha) = a'$. (This follows, since $b(x, y) = 1$, only when $x = y = 1$; $e'(x, y) = 1$ when x or $y = 1$; a is always 0; and a' is always 1.) Obvious applications of these theorem-equations will prove closure for $F(a, b)$, $F(a', b)$, $F(a, e')$, and $F(a', e')$.

Now we have only $F(b, e')$ to investigate. But, if α is either of these functions, by their definitions $\alpha(1, 1) = 1$ and $\alpha(0, 0) = 0$; that is, $F(b, e')$ is at most the intersection of F_1 and F_2. But these functions are exactly b, f, g, e', which are already known to be in $F(b, e')$. Therefore, $F(b, e')$ is closed.

This completes the verification of the fields for the 21 pairs of non-redundant, insufficient function pairs listed in Table VIII(16).

K. If we consider, now, complete sets of three undefined functions, the field of investigation does not broaden immediately, as one might think. The condition of non-redundancy forces us to reject a triad α, β, γ, which contains a pair which forms either a redundant or a complete set; for in the first case, the triad would reduce, insofar as defining power is concerned, to a pair; and in the second case, the triad would be redundant, though no pair in it need be, that is, the admissible triads will be only those such that the three pairs which it is possible to select from them be all non-redundant and insufficient, that is, be all in Table VIII.

If we examine Table VIII in this way, we find a total of only ten non-redundant triads (which establishes, incidentally, that of the $C_{16,3}$, that is, 560 possible triads, only ten are non-redundant) and determine, as in §§E and F, the fields for each of these ten triads. Of the ten triads, six were found to be sufficient, that is, we discovered six complete sets of three undefined functions, as follows:

TABLE X

a, b, h	b, h, h'	h, e', h'
a, h, e'	b, a', h'	a', e', h'

This listing is believed to be new, and answers completely the question: What non-redundant triads of undefined functions can define the calculus of sentences?

We know, from the discussion just above, that these triads are non-redun-

(16) Of the fields listed in Table VIII, only 10 were distinct: F_i ($i = 1, 2, 3, 4, 5$), and the fields for the 5 pairs last investigated.

dant and we can quickly show that they are sufficient, by the method used in §I for pairs of functions. We give (as in Table VII for pairs) the definitions, in terms of each triad, of the function e, or b', either of which can generate C.

TABLE XI

$a, b, h;$	$b' = h(b, a)$
$a, h, e';$	$e = h(e', a)$
$b, h, h';$	$a = h'(f, f);\ b' = h(b, a)$
$b, a', h';$	$b' = h'(a', b)$
$h, e', h';$	$a = h'(f, f);\ e = h(e', a)$
$a', e', h';$	$e = h'(a', e')$

L. There remain, now, only four non-redundant triads which are still not sufficient. We list them with their fields below:

TABLE XII

$F(a, b, a') = a, f, g; b; a'$
$F(a, b, e') = a, f, g; b; e'$
$F(a, a', e') = a, f, g; a'; e'$
$F(b, a', e') = b, f, g; a'; e'$

Since $F(\alpha, \beta, \gamma)$ contains at least $f, g, \alpha, \beta, \gamma$, the fields above contain at least the functions listed. To prove the fields closed, consider the following field F_6.

F_6: a, b, f, g, a', e', and the familiar theorem-equations

$$\begin{aligned} a(\beta, \gamma) &= a, & a'(\beta, \gamma) &= a', \\ b(a, \beta) &= a, & e'(a, \beta) &= \beta, \\ b(a', \gamma) &= \gamma, & e'(a', \gamma) &= a', \\ b(\beta, \beta) &= \beta, & e'(\beta, \beta) &= \beta, \\ b(\beta, \gamma) &= b(\gamma, \beta), & e'(\beta, \gamma) &= e'(\gamma, \beta). \end{aligned}$$

Since β, γ are assumed to be in F_6, the only functions to investigate are those possibly not in $F(b)$ or in $F(e')$; that is, $b(b, e')$ and $e'(b, e')$, which turn out to be respectively b and e'. Obvious applications of these theorems to the triads in Table XII will prove the fields closed as given(16).

Notice that the four non-redundant, insufficient triads are the four triads which can be selected from the tetrad a, b, a', e'.

(16) These fields are distinct, and are the last fields we shall list. We may indicate here the different fields we have come across in this investigation. The field C is of course generated by every sufficient set. In Table III, there were 11 different fields including C. In Table IX, we found 10 different fields, and now just 4 more. We have found, then, 25 fields generated by binary functions, out of the 66 different fields that Post states can be generated by functions of fewer than 4 arguments.

We have already found complete sets of one, two, and three functions. If a complete set of four undefined functions were to exist, then the condition of non-redundancy demands that the four possible triads which might be selected from it be all non-redundant and insufficient, i.e., that these four possible triads be precisely those of Table XII, that is, if any tetrad is to form a complete set, it must possess at least the properties which are enjoyed only by the tetrad a, b, a', e'. It must also be capable of defining all functions in C. But it is evident from the discussion above that $F(a, b, a', e')$ is exactly F_6, that is, $F(a, b, a', e') = a, b, f, g, a', e'$. We have shown, then, that the only non-redundant tetrad is insufficient, and therefore, that there can be no complete tetrad.

To conclude this investigation, we remark that though it is possible to add another function to this tetrad to make it a sufficient quintad, this quintad would be necessarily redundant, since the five possible tetrads in the quintad should all be non-redundant, and there is just one such tetrad: a, b, a', e'. A fortiori, there exists no complete set of more than five undefined functions.

M. We may summarize our results and translate them into the more familiar language of the calculus of sentences:

1. There are just two complete sets of single functions: e; b'; which are Sheffer's stroke function and its dual,—*joint rejection* and *alternate rejection*.

2. There are just 9 complete sets of two functions—those named in Table VI:

negation, implication;	*negation, non-implication*;
negation, conjunction;	*negation, disjunction*;
falsity, implication;	*non-equivalence, implication*;
verity, non-implication;	*equivalence, non-implication*;

implication, non-implication.

3. There are just six complete sets of three functions—those named in Table X, or, in translation:

falsity, conjunction, equivalence;
falsity, disjunction, equivalence;
verity, conjunction, non-equivalence;
verity, disjunction, non-equivalence;
conjunction, equivalence, non-equivalence;
disjunction, equivalence, non-equivalence.

4. There is no complete set of more than three functions[17].

NEW YORK UNIVERSITY,
NEW YORK, N. Y.

---

[17] The referee points out that the results are true (in a certain sense) in any Boolean algebra, and not merely in the two-valued calculus of propositions. "That is, the only complete sets of Boolean functions of two variables *without parameters* are the sets he gives."

ON DISTORTION IN PSEUDO-CONFORMAL MAPPING

BY

STEFAN BERGMAN AND D. C. SPENCER

1. Introduction

1.1. Suppose that the functions $w_k(z_1, z_2)$, $k=1, 2$(1), are regular in a four-dimensional domain $\mathfrak{B}_1$ of the complex variables z_1, z_2. The transformation $\mathbf{w}$ of $\mathfrak{B}_1$ into a domain $\mathfrak{B}_2$ by a pair of functions w_k, for which $\partial(w_1, w_2)/\partial(z_1, z_2)$ does not vanish identically, is called a PT (pseudo-conformal transformation). We are here concerned with general PT's in which the mapping is not necessarily one-one with respect to the schlicht space of the variables. Suppose that $\mathfrak{B}_1$ is a fixed schlicht domain in the space (z_1, z_2), which contains the point $(0, 0)$ in its interior. Let

$$\mathcal{C} = \mathcal{C}(\mathfrak{B}_1)$$

be some specified subfamily of the family of domains $\mathfrak{B}_2$ obtainable from $\mathfrak{B}_1$ by PT's $\mathbf{w}$ which satisfy a certain normalization hypothesis at $(0, 0)$; and let

$$M = M(\mathfrak{B}_2)$$

be the euclidean measure of a geometrical object defined for each $\mathfrak{B}_2$ of $\mathcal{C}$(2). The problem of the bounds between which $M(\mathfrak{B}_2)$ may vary for $\mathfrak{B}_2 \in \mathcal{C}(\mathfrak{B}_1)$ forms an important chapter in the theory(3).

In this paper we give inequalities for the volume of a subdomain $\mathfrak{B}_2^\dagger$ of $\mathfrak{B}_2$, where

$$\mathfrak{B}_2^\dagger = S(\mathfrak{B}_2)$$

is derived from $\mathfrak{B}_2$ by the following geometrical operation. We cut $\mathfrak{B}_2$ with the plane $\arg w_k = \psi_k = \text{constant}$, the intersection $\mathfrak{S}_2^2(\psi_1, \psi_2)$ being a plane set which consists in general of many disconnected components. We suppose that $\mathbf{w}(0, 0) = (0, 0)$; we denote by $\mathfrak{S}_2^{\dagger 2}(\psi_1, \psi_2)$ that connected component of the intersection of $\mathfrak{B}_2$ with $\arg w_k = \psi_k$ for which $w(0, 0)$ is a boundary point(4), and define $\mathfrak{B}_2^\dagger = \mathbf{S}(\mathfrak{B}_2)$ to be the domain of which the intersection with every

Presented to the Society, September 11, 1940; received by the editors November 25, 1940.

(1) To avoid endless repetition, we shall omit the statement "$k=1, 2$" hereafter. Whenever the index k occurs, it is to have the values 1 and 2.

(2) For example, M may be the distance of the boundary of $\mathfrak{B}_2$ from the origin, or the volume of $\mathfrak{B}_2$.

(3) As it does in conformal mapping. See also Bergman [4, chap. 5] and [6, pp. 3–7]. The numbers in brackets refer to the bibliography, p. 162.

(4) Since in general the domain $\mathfrak{B}_2$ is not schlicht, there may exist points $(0, 0)$ lying in other sheets which are images of points (z_1, z_2) different from $(0, 0)$. For a more detailed description of $\mathfrak{S}_2^{\dagger 2}(\psi_1, \psi_2)$ see §7.

plane $\arg w_k = \psi_k$, $0 \leqq \psi_k \leqq 2\pi$, is (with exception of a set of measure zero) $\mathfrak{S}_2^{\dagger 2}(\psi_1, \psi_2)$. (Figure 1 is a schematic diagram in which the quadrant $\arg w_k = \psi_k$ is replaced by the half-line $\arg w = \psi$, and represents the construction in one variable.) We obtain an inequality for the volume of $\mathfrak{B}_2^{\dagger}$ which depends: (i) on the initial domain $\mathfrak{B}_1$(5); (ii) on an average over $\mathfrak{B}_1$ of the

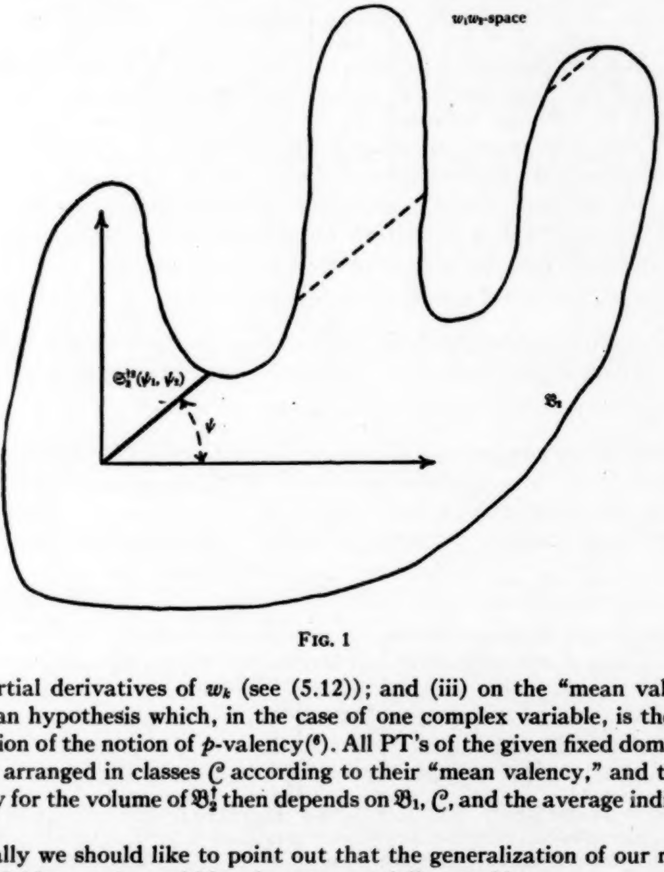


FIG. 1

first partial derivatives of w_k (see (5.12)); and (iii) on the "mean valency" of $\mathfrak{B}_2^{\dagger}$—an hypothesis which, in the case of one complex variable, is the generalization of the notion of p-valency(6). All PT's of the given fixed domain $\mathfrak{B}_1$ may be arranged in classes $\mathcal{C}$ according to their "mean valency," and the inequality for the volume of $\mathfrak{B}_2^{\dagger}$ then depends on $\mathfrak{B}_1$, $\mathcal{C}$, and the average indicated in (ii).

Finally we should like to point out that the generalization of our results to n variables, $n > 2$, would involve no essentially new ideas.

(5) We take for $\mathfrak{B}_1$ the unit bicylinder, but an inequality analogous to the one given in this paper can be found for any domain $\mathfrak{B}_1$, and by the same general methods. However, the hypotheses of the theorem would have to be changed accordingly. Concerning other distortion theorems, see Bergman [4, pp. 124–158] and [5, chap. 9].

(6) The idea of mean valency in one variable was first introduced by Spencer in his paper [1].

2. Notation

2.1. Manifolds will be denoted by German letters, the superscript denoting the dimension of the manifold. However, we omit the superscript if the manifold is four-dimensional or a point. In operations on sets we use the customary symbols[7]. For example, $\mathfrak{E}_1^n \cdot \mathfrak{E}_2^m$ denotes the common part (intersection) of $\mathfrak{E}_1^n$ and $\mathfrak{E}_2^m$, $\mathfrak{E}_1^m \times \mathfrak{E}_2^n$ the topological product of the two sets, $\mathfrak{a}^n \in \mathfrak{E}^m$ means that "$\mathfrak{a}^n$ is contained in $\mathfrak{E}^m$," and so on.

We use the symbol $\mathfrak{S}$ to denote logical summation over a set,—for example $\mathfrak{S}_{\alpha \varepsilon \mathfrak{N}^m} \mathfrak{E}^n(\alpha)$ is the sumset of the family of sets $\mathfrak{E}^n(\alpha)$ where α, the parameter of the family, runs through a set $\mathfrak{N}^m$. By $\mathrm{E}[\cdots]$ we mean the set of points the coordinates of which satisfy the conditions indicated in the brackets. However, the intersection of a manifold $\mathfrak{E}^n$ with $\mathrm{E}[\cdots]$ will be denoted $\mathfrak{E}^n \cdot [\cdots]$. By $\mathrm{V}(\mathfrak{B})$ we denote the volume of a domain $\mathfrak{B}$, by $\mathrm{A}(\mathfrak{B}^2)$ the area of $\mathfrak{B}^2$.

In this paper we shall consider the PT's of manifolds of the z_1z_2-space, $z_k = x_k + iy_k$, into manifolds of the w_1w_2-space, $w_k = u_k + iv_k$, where x_1, y_1, x_2, y_2, and u_1, v_1, u_2, v_2, are the Cartesian coordinates of the respective spaces. The element of volume in the z_1z_2-space will be denoted by $d\omega_z$, in the w_1w_2-space by $d\omega_w$. Manifolds of the z_1z_2-space will be denoted by (German) letters with subscript 1, and those of the w_1w_2-space will be denoted by letters with subscript 2. Let $C(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ be some coordinate system in four-dimensional space. Given a manifold $\mathfrak{M}^n$, we define the projection of $\mathfrak{M}^n$ on $C_1(\alpha_p, \cdots, \alpha_q)$, $\mathrm{P}_{\alpha_p \cdots \alpha_q}(\mathfrak{M}^n)$ say, as the totality of number sets $(A_p, \cdots, A_q)$ for which (A_1, A_2, A_3, A_4) are the coordinates of some point of $\mathfrak{M}^n$.

Finally we shall say that a function $f(z_1, z_2)$ is regular in $\mathfrak{M}^n$, $n \leqq 3$, if there exists a four-dimensional manifold containing $\mathfrak{M}^n$ in which f is regular. In particular we shall say that a function $f(\sigma_1, \sigma_2)$ is an analytic function of two real variables σ_1, σ_2, regular in a domain $\mathfrak{F}^2$, if it can be developed in a power series $\sum a_{mn}(\sigma_1 - \sigma_1^{(0)})^m(\sigma_2 - \sigma_2^{(0)})^n$ at any point $(\sigma_1^{(0)}, \sigma_2^{(0)})$ of $\mathfrak{F}^2$, the series converging absolutely and uniformly in some neighborhood of $(\sigma_1^{(0)}, \sigma_2^{(0)})$.

3. Results for functions of one variable

3.1. The result which we prove is a generalization to two complex variables of a theorem of Golusin [1], which in turn is a generalization of the lemma of Schwarz.

Suppose that $w(z)$ is regular for $|z| < 1$, that $w(0) = 0$, $|w'(0)| = 1$, and that $\mathfrak{B}_2^2$ is the map of $\mathrm{E}[|z| < 1]$ by $\mathbf{w} = [w = w(z)]$. Then $\mathrm{A}(\mathfrak{B}_2^2)$ (multiply-covered regions being counted multiply) is not less than π. Golusin's theorem states that not only $\mathrm{A}(\mathfrak{B}_2^2)$, but also $\mathrm{A}(\mathfrak{B}_2^{\dagger 2})$, is at least equal to π, where the domain $\mathfrak{B}_2^{\dagger 2}$ is the two-dimensional analogue of the domain $\mathfrak{B}^\dagger$ defined above (see Figure 1). A more precise version of Golusin's theorem was given by Bermant [1] and lately a still more precise one by Spencer [2].

(7) See for example Hausdorff, *Mengenlehre*, 2d edition, 1928, Section 1.

The final form of the results may be stated as follows:

Suppose that $w(z)$ is regular for $\mathrm{E}[|z|<1]$, *that* $w(0)=0$, $|w'(0)|=1$, *and let* $\mathfrak{B}_2^{\dagger 2}$ *be the portion of* $\mathfrak{B}_2^2$ *which is visible to an eye placed at* $w(0)$ *if all boundary elements of* $\mathfrak{B}^2$ *are opaque. Then*

$$\mathrm{A}[\mathbf{w}^{-1}(\mathfrak{B}_2^{\dagger 2})]\cdot \mathrm{A}(\mathfrak{B}_2^{\dagger 2}) \geqq \pi^2, \tag{3.1.1}$$

with equality only if $w=z$[8].

Now the generalization of these results to two variables is not immediate; we require certain new ideas and methods, in particular the following two concepts: (i) Four-dimensional "mean valency," and a new representation for the volume of a four-dimensional (in general multiply-covered) domain (see §4). Mean valency here assumes an interesting form, which puts the general concept of valency in a new light, even in one variable. (ii) A new kind of area, "*B*-area." If, in passing from conformal transformations to PT's, ordinary area is replaced by volume, length of curves by *B*-area of surfaces, a generalization is obtained, involving *B*-area and volume, of inequalities between length and area (see Bergman [4, Chap 5, §6, pp. 147–149]), and this method is adopted in §8.

4. Mean valency

4.1. Given a space $\mathfrak{U}^{2n}=\mathrm{E}[|w_{\mathrm{H}}|<\infty, \mathrm{H}=1, 2, \cdots, n]$ we first choose a suitable coordinate system $\mathfrak{W}$ in $\mathfrak{U}^{2n}$, and then some particular one of the coordinates, $\lambda=\lambda(w)$, say. If $\mathfrak{B}^{2n}$ is an arbitrary (in general multiply-covered) domain, we denote by $\mathfrak{B}^{2n}(\Lambda)$ the domain $\mathfrak{B}^{2n}\cdot[\lambda<\Lambda]$, that is, the subdomain of $\mathfrak{B}^{2n}$ for which $0\leqq\lambda<\Lambda$. Suppose now that $\mathfrak{B}_0^{2n}$ is a given domain. Then we say that $\mathfrak{B}^{2n}$ is mean p-valent with respect to $\mathfrak{B}_0^{2n}$ in the direction λ if

$$\mathrm{V}[\mathfrak{B}^{2n}(\Lambda)] \leqq p\mathrm{V}[\mathfrak{B}_0^{2n}(\Lambda)] \tag{4.1.1}$$

for all $\Lambda>0$, where p is a positive real number[9]. Since in this paper we shall be concerned with only one domain $\mathfrak{B}_0^{2n}$, and with fixed direction of measure λ in a fixed system of coordinates, we shall say simply that $\mathfrak{B}^{2n}$ is mean p-valent. For example, suppose that $n=1$, and that we choose for $\mathfrak{W}$ the polar co-

(8) $\mathbf{w}^{-1}(\mathfrak{B}_2^{\dagger 2})$ and $\mathfrak{B}_2^{\dagger 2}$ are both schlicht domains which contain the points $z=0$, $w=0$, respectively (but in two variables $\mathfrak{B}_2^{\dagger}$ is not necessarily schlicht). $\mathfrak{B}_2^{\dagger 2}$ is a "star-like" domain (the "star" of $\mathfrak{B}_2^2$ with respect to the point $w(0)$). Any point of $\mathfrak{B}_2^{\dagger 2}$ can be reached by starting at $w(0)=0$ and travelling outward along a radius.

If $\mathfrak{B}^2$ is itself a "star" domain, then $\mathfrak{B}^{\dagger 2}=\mathfrak{B}^2$; the difference between the areas of $\mathfrak{B}^2$ and $\mathfrak{B}^{\dagger 2}$ is in a certain sense inversely as the "star-likeness" of $\mathfrak{B}^2$.

(9) We could, of course, define mean valency without introducing the domain $\mathfrak{B}_0^{2n}$, merely replacing the right-hand member of (4.1.1) by a function of Λ; but the geometrical interpretation seems to us desirable here.

When $n=1$, V becomes area, which we denote by A.

ordinates λ, ψ(10), so that $w=\lambda e^{i\psi}$. If then we take $\mathfrak{B}_0^2$ to be the whole plane, that is

$$(4.1.2)\qquad \mathfrak{B}_0^2 = \mathrm{E}[\,|\,w\,|\leqq 1] + \mathrm{E}[\,|\,w\,|\geqq 1],$$

the inequality (4.1.1) becomes the hypothesis of mean p-valency in one variable (see Spencer [1]).

In the case of $\mathfrak{U}^4$ we write $w_k=r_ke^{i\psi_k}$, and define λ and μ by the equation

$$(4.1.3)\qquad \mu + i\lambda = (r_1 + ir_2)^2.$$

We take now for $\mathfrak{W}$ the coordinates(11) λ, μ, ψ_1, ψ_2. Next, in analogy with (4.1.2), we take

$$(4.1.4)\qquad \mathfrak{B}_0 = \mathrm{E}[\,|\,w_k\,|\leqq 1] + \mathrm{E}[\,|\,w_k\,|\geqq 1].$$

The coordinate system λ, μ, ψ_1, ψ_2, and standard domain $\mathfrak{B}_0$ defined by (4.1.3) and (4.1.4), respectively, will be used throughout this paper.

4.2. We now formulate these definitions analytically, and, for purposes of comparison, we first formulate the hypothesis in one variable.

Let(12) $\mathrm{n}(\lambda,\psi)=\mathrm{n}_{\mathfrak{B}^2}(\lambda,\psi)$ be the number of times $\mathfrak{B}^2$ covers the point (λ,ψ) of the schlicht space. We define

$$(4.2.1)\qquad \mathrm{p}_{\mathfrak{B}^2}(\lambda) = (1/2\pi)\int_{-\pi}^{\pi}\mathrm{n}(\lambda,\psi)d\psi.$$

Then

$$(4.2.2)\qquad \mathrm{A}[\mathfrak{B}^2(\Lambda)] = \int_0^{\Lambda}\mathrm{p}(\lambda)d(\pi\lambda^2).$$

In particular, taking $\mathfrak{B}^2=\mathfrak{B}_0^2$, we have $\mathrm{p}(\lambda)=\mathrm{p}_{\mathfrak{B}_0^2}(\lambda)=1$ for all λ, and so

$$(4.2.3)\qquad \mathrm{A}[\mathfrak{B}_0^2(\Lambda)] = \int_0^{\Lambda}d(\pi\lambda^2).$$

Substituting from (4.2.3) and (4.2.2) into (4.1.1), the hypothesis of mean p-valency in one variable may be expressed in the form

$$(4.2.4)\qquad \int_0^{\Lambda}\mathrm{p}(\lambda)d(\pi\lambda^2) \leqq p\int_0^{\Lambda}d(\pi\lambda^2) = p\pi\Lambda^2 \qquad (\Lambda > 0).$$

4.3. In the space of two variables we choose the coordinate system defined above by (4.1.3), and take for $\mathfrak{B}_0$ the domain (4.1.4). We let $\mathrm{n}(\lambda,\mu,\psi_1,\psi_2)$ $=\mathrm{n}_{\mathfrak{B}}(\lambda,\mu,\psi_1,\psi_2)$ be the number of times that $\mathfrak{B}$ covers the point $(\lambda,\mu,\psi_1,\psi_2)$ of the schlicht space, and define

(10) λ will be used throughout to denote the coordinate which occurs above.

(11) These coordinates assume for some purposes a role similar to that of the polar coordinates λ, ψ in one variable. See Bergman [3, §3].

(12) We omit the subscript $\mathfrak{B}^2$ when it is clear on what domain the functions n, p, $\cdots$ depend.

$$N_{\mathfrak{B}}(\lambda, \psi_1, \psi_2) = (1/4) \int_{-\infty}^{\infty} n(\lambda, \mu, \psi_1, \psi_2)(\lambda^2 + \mu^2)^{-1/2} d\mu, \tag{4.3.1}$$

$$p_{\mathfrak{B}}(\lambda) = (1/4\pi^2) \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} N_{\mathfrak{B}}(\lambda, \psi_1, \psi_2) d\psi_1 d\psi_2. \tag{4.3.2}$$

Then

$$V[\mathfrak{B}(\Lambda)] = \int_0^{\Lambda} p(\lambda) d(\pi^2\lambda^2). \tag{4.3.3}$$

In particular, if we take $\mathfrak{B} = \mathfrak{B}_0$ (cf. (4.1.4)), we have $p(\lambda) = p_{\mathfrak{B}_0}(\lambda) = (1/2)|\lg(2/\lambda)|$, and so

$$V[\mathfrak{B}_0(\Lambda)] = (p/2) \int_0^{\Lambda} |\lg(2/\lambda)| d(\pi^2\lambda^2). \tag{4.3.4}$$

Substituting from (4.3.3) and (4.3.4) into (4.1.1), the hypothesis of mean p-valency in two variables becomes

$$\int_0^{\Lambda} p(\lambda) d(\pi^2\lambda^2) \leqq (p/2) \int_0^{\Lambda} |\lg(2/\lambda)| d(\pi^2\lambda^2) \qquad (\Lambda > 0). \tag{4.3.5}$$

REMARK. If P is a point for which $\lambda = 0$, we shall (for convenience) denote the coordinates of P by $(0, \mu, \psi_1, \psi_2)$ even though ψ_1 or ψ_2 may be meaningless. For example, if $\mu > 0$, we shall write $n(0, \mu, \psi_1, \psi_2)$ for $n(0, \mu, \psi_1)$.

4.4. A necessary and sufficient condition that p-valency should imply mean p-valency with respect to a domain $\mathfrak{B}_0$ is that

$$\mathfrak{B}_0^{2n} = \mathfrak{U}^{2n}.$$

The particular definition of mean p-valency (bicylindrical mean p-valency) selected above satisfies this condition in one, but not in two, variables. Since this definition carries with it certain important metrical properties, it is natural to expect that many theorems associated with valency in one variable will be true in two variables under a hypothesis of mean valency, but false under one of valency alone.

4.5. In this paper we shall be concerned with the hypothesis only when $p = 1$. Suppose that $F(\Lambda)$ is a real function of Λ, defined for $\Lambda \geqq 0$, which satisfies the following two conditions:

$$0 \leqq \Lambda F(\Lambda) \leqq 1 \qquad (\Lambda > 0), \tag{4.5.1}$$

$$\int_0^{\infty} F(\Lambda) d\Lambda \leqq 1/2. \tag{4.5.2}$$

We shall say that a domain $\mathfrak{B}$ is mean one-valent with excess α, or that $\mathfrak{B}$ belongs to the class $\mathfrak{T}(\alpha)$, $\alpha \geqq 0$, if there exists an $F(\Lambda)$ satisfying (4.5.1) and (4.5.2) and such that, for all $\Lambda > 0$,

$$\int_0^{\Lambda} p_{\mathfrak{B}}(\lambda) d(\pi^2\lambda^2) \leqq (1/2) \int_0^{\Lambda} |\lg(2/\lambda)| d(\pi^2\lambda^2) + \alpha\Lambda^3 F(\Lambda). \tag{4.5.3}$$

5. Statement of the theorem

5.1. We write

$$z_k = \rho_k e^{i\theta_k}, \qquad d\omega_z = \rho_1\rho_2 d\rho_1 d\rho_2 d\theta_1 d\theta_2,$$

and denote the unit bicylinder $E[|z_k| < 1]$ by $\mathfrak{B}_1$. By $\mathfrak{B}^\dagger$ we mean the domain obtained from a domain $\mathfrak{B}$ by the geometrical operation discussed in the introduction (§1)([13]).

Theorem. *Suppose that the* PT

$$(5.1.1) \qquad \mathbf{w}: \ w_k = w_k(z_1, z_2) \equiv \overset{(k)}{a}_{2-k,k-1} z_k + z_1^{2-k} z_2^{k} f_k(z_1, z_2),$$

where $|a^{(k)}_{2-k,k-1}| = 1$, *and where* $f_k(z_1, z_2)$ *is regular in* $\mathfrak{B}_1$, *maps* $\mathfrak{B}_1$ *on a domain* $\mathfrak{B}_2$. *Then* (i) *if* $\mathfrak{B}_2^\dagger \in \mathfrak{T}(\alpha_1)$ [*in particular if* $\mathfrak{B}_2 \in \mathfrak{T}(\alpha_1)$], *and* (ii) *if*

$$(5.1.2) \qquad \iiiint_{\mathbf{w}^{-1}(\mathfrak{B}_2^\dagger)} \sum_{k=1}^{2} \left[\frac{1}{|z_k|^2} \left| \frac{\partial w_1}{\partial z_{3-k}} \cdot \frac{\partial w_2}{\partial z_{3-k}} \right| \right] \frac{d\omega_z}{|w_1 w_2|} \leqq \alpha_2$$

(*in particular if* $\iiiint_{\mathfrak{B}_1} \cdots \leqq \alpha_2$), *we have*

$$(5.1.3) \qquad \mathrm{V}[\mathbf{w}^{-1}(\mathfrak{B}_2^\dagger)] \cdot \mathrm{V}(\mathfrak{B}_2^\dagger) \geqq K(\alpha_1, \alpha_2)$$

where

$$(5.1.4) \qquad K(\alpha_1, \alpha_2) = (\pi/4)^4 \exp\left[- \frac{4\alpha_1 + \alpha_2}{\pi^2 \lg 2} \right].$$

Remark. We prove the theorem subject to the additional hypothesis that([14]) $\mathrm{n}_{\mathfrak{B}_2}(0, \mu, \psi_1, \psi_2) \leqq 1$, but indicate in §9 how this restriction may be removed. However, since the removal of the condition involves only technical complications and no new ideas, we give only a sketch.

5.2. Let $\mathfrak{C}$ be the class of domains $\mathfrak{B}_2$ arising from the bicylinder([15]) $\mathfrak{B}_1$ by PT's $\mathbf{w}$ normalized at (0, 0). If P is a point of $\mathfrak{B}_2$, we may take for the coordinates of P the coordinates of the point $P' = \mathbf{w}^{-1}(P)$, and thus define in $\mathfrak{B}_2$ a new system of coordinates $\mathfrak{K}(\mathfrak{B}_2)$, say. The intersection $\mathfrak{S}_2^{12}(\psi_1, \psi_2)$ defined in §1 has the property that its area (multiply-covered portions being counted multiply) is the same whether measured in the A- or B-sense (A is ordinary area; for the definition of B-area see Bergman [2], or §8.1). The integral on the left-hand side of (5.1.2) is a measure of the invariance of this property to the change of coordinates, and (5.1.2) states that the average with respect to (ψ_1, ψ_2) of the difference between the A- and B-areas of $\mathfrak{S}_2^{12}(\psi_1, \psi_2)$ in the new system $\mathfrak{K}(\mathfrak{B}_2)$ does not exceed α_2.

(13) For a more detailed description of $\mathfrak{B}^\dagger$ we refer the reader to §7.

(14) For the definition of n see §4.

(15) The bicylinder $\mathfrak{B}_1$ is the "representative domain" of $\mathfrak{C}$ with respect to (0,0) see Bergman [4, chap. 7, pp. 215–230] and [5, chap. 10]).

5.3. By the theorem in one variable cited in §3, we see that the result of the theorem is true with $K=\pi^4$ when $\mathbf{w}$ is a product PT. In the above theorem we suppose that all but one of the coefficients which border the matrices of coefficients of w_1 and w_2 are zero, and this hypothesis in itself([16]) excludes from consideration all product PT's except the trivial rotation $w_k = a^{(k)}_{2-k,k-1} z_k$. Although product PT's (in which *only* the coefficients bordering the matrices of w_1 and w_2 appear) are a small class compared to the class of PT's admitted by our hypothesis (even for fixed α_1, α_2), nevertheless the question arises as to why product PT's are not a special case of our theorem. The answer is that the inclusion of "product-like" PT's would involve nothing essentially new, but would require two mutually exclusive sets of hypotheses in our theorem.

6. Preliminary remarks concerning the proof

6.1. The proof is divided into two parts—a geometrical part (§7), and an analytical part (§8). In order not to interrupt the argument later, we sum up here certain facts to which we shall constantly refer.

6.2. First, we may assume that the functions w_k are regular in the closed bicylinder $\mathfrak{D}_1$ ($\mathfrak{D}_1$ is the closure of $\mathfrak{B}_1$); therefore in some domain which contains $\mathfrak{D}_1$. For, having proved the theorem in this special case, we can apply it to the functions

$$\rho^{-1} w_k(\rho z_1, \rho z_2) = a^{(k)}_{2-k,k-1} z_k + z_1^{3-k} \cdot z_2^{k} \rho^{2} f_k(\rho z_1, \rho z_2) \tag{6.2.1}$$

where $\rho < 1$, and then let $\rho \to 1$. In particular, we may assume that f_1 and f_2 are bounded in $\mathfrak{D}_1$.

Secondly, from the hypothesis that the PT is schlicht on $|z_k| = 0$, it now follows that there exists a δ such that the mapping is schlicht for $|z_1 z_2| < \delta$. Assuming, therefore, that ϵ is sufficiently small, we see that the pair of inverse functions $z_k(w_1, w_2)$ have the form

$$z_k(w_1, w_2) = w_k / a^{(k)}_{2-k,k-1} + w_1^{3-k} w_2^{k} g_k(w_1, w_2), \tag{6.2.2}$$

where the g_k are both regular in the portion of $\mathfrak{D}_2$ ($\mathfrak{D}_2$ is the closure of $\mathfrak{B}_2$) for which either $|w_1| < \epsilon$, or $|w_2| < \epsilon$. A particular consequence of (6.2.2) is that, if $\lambda \leqq \lambda_0$, points (w_1, w_2) of $\mathfrak{D}_2$ for which $|w_1 w_2| = \lambda$, are transformed into points (z_1, z_2) which satisfy

$$\begin{aligned} |z_1 z_2| &= |w_1 + a^{(1)}_{10} w_1^{2} w_2 g_1| \cdot |w_2 + a^{(2)}_{01} w_1 w_2^{2} g_2| \\ &= |w_1 w_2| + O(|w_1 w_2|^2) = \lambda + O(\lambda^2), \end{aligned} \tag{6.2.3}$$

and conversely.

([16]) Furthermore, product PT's do not in general satisfy the mean-valency restrictions of our theorem when α_1 is finite (but they satisfy (5.1.2) with $\alpha_2 = 0$).

7. PROOF. GEOMETRICAL PART

7.1. Throughout this section we assume that the PT's which we consider satisfy not only the hypotheses of the theorem, but also the additional hypotheses of §6. In particular, therefore, we assume that the functions w_k are regular for $E[|z_k| \leq P_k]$, where $P_k > 1$. Such a PT transforms the (closed) bicylinder

$$\mathfrak{D}_1(P_k) = E[\,|z_k| \leq P_k]$$

into a domain $\mathfrak{D}_2(P_k)$ imbedded in a many-sheeted Riemannian space. At each point P of $\mathfrak{B}_1(P_k) = E[\,|z_k| < P_k]$ at which

$$J(z_1, z_2) = \partial(w_1, w_2)/\partial(z_1, z_2) \neq 0 \tag{7.1.1}$$

the PT is locally schlicht: that is to say, the different sheets are not connected in a neighborhood of $\mathbf{w}(P)$. The surface $\mathfrak{B}_2^2 = \mathbf{w}(\mathfrak{B}_1^2)$, where

$$\mathfrak{B}_1^2 = E[J(z_1, z_2) = 0], \tag{7.1.2}$$

is the "branch surface" of the Riemannian space in which $\mathfrak{D}_2(P_k)$ lies. We divide the points of $\mathfrak{B}_2^2$ into two classes: (i) branch points of the first kind, which are transforms of points (z_1, z_2) at which $J(z_1, z_2)$ has only *one* prime factor (which may, however, be raised to a power higher than the first); and (ii) branch points of the second kind, which correspond to points (z_1, z_2) where $J(z_1, z_2)$ has more than one prime factor ([17]). Since J has only a finite number of factors $(z_2 - a_2^{(p)})$, we can, by the Weierstrass preparation theorem and the Heine-Borel covering theorem, cover $\mathfrak{B}_1(P_k)$ with finitely many domains $\mathfrak{G}_{1,j}$, $j = 1, 2, \cdots, n$, such that([18]), in each $\mathfrak{G}_{1,j}$

$$J(z_1, z_2) = z_2^{\mu} \prod_{H} [z_1 - \alpha_H(z_2)]^{\lambda_H} \Omega(z_1, z_2)$$

where $\Omega(z_1, z_2)$ is regular and nonvanishing in $\mathfrak{G}_{1,p}$, and where the $\alpha_H(z_2)$ are algebroid functions([19]).

The function

$$\Delta(z_2) = \prod_{H,\rho;\, H \neq \rho} [\alpha_H(z_2) - \alpha_\rho(z_2)]^2,$$

being a regular function of z_2, vanishes at only a finite number of points $a_2^{(p)}$. Since there are only finitely many points $(a_1^{(p)}, a_2^{(p)})$, $a_1^{(p)} = \alpha_H(a_2^{(p)})$ there are

([17]) See Osgood [1, chap. 2, §20, p. 111]. With exception of finitely many points $(z_1^{(\nu)}, z_2^{(\nu)})'$ we can suppose that these factors have the form $[z_1 - \alpha(z_2)]$.

([18]) For the sake of brevity we suppose that the point in the neighborhood of which we consider $J(z_1, z_2)$ is the origin.

([19]) A function $\alpha = \alpha(z)$ satisfying the equation $\alpha^m + g_1(z)\alpha^{m-1} + \cdots + g_m(z) = 0$, where $g_p(z)$, $p = 1, 2, \cdots, m$, are analytic functions of one complex variable regular in $\mathfrak{N}^2$, is said to be "algebroid in $\mathfrak{N}^2$."

only finitely many branch points of the second kind and we can, therefore, with finitely many exceptions, write in the neighborhood of every point $(z_1^{(0)}, z_2^{(0)})$ either

$$J(z_1, z_2) = [z_1' - \alpha(z_2')]^{\mu}\Omega(z_1', z_2'),$$

or

$$J(z_1, z_2) = z_2'^{\mu}\Omega(z_1', z_2'),$$

where $z_k' = z_k - z_k^{(0)}$, and where $\alpha(z_2')$ is a regular function of z_2. We remove from $\mathfrak{B}_2^2$ the branch points of the second kind, and denote the remaining surface by $\mathfrak{C}_2^2$.

By the Heine-Borel theorem and the results of Osgood [1] we can cover every closed subdomain of $\mathfrak{C}_2^2$ with finitely many neighborhoods $\mathfrak{F}_{2,p}$ such that to each $\mathfrak{F}_{2,p}$ there corresponds a uniformizing transformation

$$\sigma_1 = (w_1 - H(w_2))^{1/m}, \qquad \sigma_2 = w_2, \tag{7.1.3}$$

or $\sigma_1 = w_1$, $\sigma_2 = (w_2 - H(w_1))^{1/m}$, where H is a suitably chosen analytic function of one complex variable, such that the PT

$$\sigma_k = \sigma_k(z_1, z_2) \tag{7.1.4}$$

is one-one in $\mathfrak{F}_{2,p}$.

7.2. We now derive certain trivial properties of $\mathbf{w}$, in particular, that $\mathbf{w}$ is a topological transformation. We write

$$\begin{aligned}\mathfrak{B}_{1,\epsilon}(\mathrm{P}_k) = \mathfrak{B}_1(\mathrm{P}_k) &- \underset{k=1}{\overset{2}{\mathfrak{S}}} \mathrm{E}[\,|z_k| \leqq \epsilon] \\ &- \underset{(\zeta_1,\zeta_2)\in\mathfrak{B}_1^3}{\mathfrak{S}} \mathrm{E}[\,|z_1 - \zeta_1|^2 + |z_2 - \zeta_2|^2 \leqq \epsilon^2],\end{aligned} \tag{7.2.1}$$

$$\mathfrak{B}_{2,\epsilon}(\mathrm{P}_k) = \mathbf{w}[\mathfrak{B}_{1,\epsilon}(\mathrm{P}_k)]. \tag{7.2.2}$$

Lemma 1. *There exists a number C, depending only on $\mathbf{w}$ and ϵ, such that we have the following inequality connecting the length* $\mathrm{L}(\mathfrak{l}_1^1)$ *of a curve* $\mathfrak{l}_1^1$ *in* $\mathfrak{B}_{1,\epsilon}(\mathrm{P}_k)$ *with the length of its transform* $\mathfrak{l}_2^1 = \mathbf{w}(\mathfrak{l}_1^1)$:

$$C^{-1}\mathrm{L}(\mathfrak{l}_2^1) \leqq \mathrm{L}(\mathfrak{l}_1^1) \leqq C\mathrm{L}(\mathfrak{l}_2^1). \tag{7.2.3}$$

In fact, let $\mathfrak{l}_2^1 = \mathrm{E}\,[w_k = w_k(u),\ 0 \leqq u \leqq 1]$, where $w_k(u)$ is differentiable in u. Then

$$\begin{aligned}\{d\mathrm{L}[\mathfrak{l}_1^1(u)]\}^2 &= \left(\sum_{j=1}^{2}\left|\frac{dz_j}{du}\right|^2\right)du^2 = \sum_{j=1}^{2}\left|\sum_{k=1}^{2}\frac{\partial z_j}{\partial w_k}\frac{\partial w_k}{\partial u}\right|^2 du^2 \\ &\leqq \left[\sum_{j=1}^{2}\sum_{k=1}^{2}\left|\frac{\partial z_k}{\partial w_j}\right|^2\right]\cdot\left[\sum_{k=1}^{2}\left|\frac{\partial w_k}{\partial u}\right|^2\right]du^2,\end{aligned}$$

and analogously

$$\{dL[l_2^1(u)]\}^2 \leqq \left[\sum_{j=1}^{2}\sum_{k=1}^{2}\left|\frac{\partial w_k}{\partial z_j}\right|^2\right]\cdot\left[\sum_{k=1}^{2}\left|\frac{\partial z_k}{\partial u}\right|^2\right]du^2.$$

Setting

$$C = \max\left[\max_{(w_1,w_2)\in\mathfrak{B}_{2,\epsilon}(P_k)}\sum_{j=1}^{2}\sum_{k=1}^{2}\left|\frac{\partial z_k}{\partial w_j}\right|^2,\ \max_{(z_1,z_2)\in\mathfrak{B}_{1,\epsilon}(P_k)}\sum_{j=1}^{2}\sum_{k=1}^{2}\left|\frac{\partial w_k}{\partial z_j}\right|^2\right],$$

we have the lemma. Furthermore $L(\mathfrak{l}_2^1)\leqq CL(\mathfrak{l}_1^1)$ throughout $\mathfrak{B}_1(P_k)$; and $L(\mathfrak{l}_1^1)\leqq CL^*(\mathfrak{l}_2^1)$ if $\mathfrak{l}_1^1\in\mathfrak{F}_{1,m}$, where $L^*(\mathfrak{l}_2^1)$ denotes the length of $\mathfrak{l}_2^1$ in the $\sigma_1\sigma_2$-space (see (7.13)), and $\mathfrak{F}_{1,m}=\mathbf{w}^{-1}(\mathfrak{F}_{2,m})$.

Since the transforms of the hyperspheres of radius ϵ and centers $\mathbf{w}^{-1}(w_1^{(\nu)}, w_2^{(\nu)})$ lie in an arbitrarily small neighborhood of $(w_1^{(\nu)}, w_2^{(\nu)})$, where $(w_1^{(\nu)}, w_2^{(\nu)})$, $\nu=1, 2, \cdots, n$, are branch points of the second kind, we see that, if a sequence of points P_H tends to a limit P_0 in the z_1z_2- (or w_1w_2-) space, then $\lim_{H\to\infty}\mathbf{w}(P_H)=\mathbf{w}(P_0)$ (or $\lim_{H\to\infty}\mathbf{w}^{-1}(P_H)=\mathbf{w}^{-1}(P_0)$). Also

$$\lim_{\epsilon\to 0} V[\mathfrak{B}_{2,\epsilon}(P_k')] = V[\mathfrak{B}_2(P_k')] \qquad (P_k' < P_k).$$

7.3. We shall denote the transform of the bicylinder $\mathfrak{B}_1(\tau_k)=E[|z_k|<\tau_k]$, $\tau_k<P_k$, by $\mathfrak{B}_2(\tau_k)$. The boundary of $\mathfrak{B}_2(\tau_k)$ consists of the two hypersurfaces[20]

$$(7.3.1)\qquad \mathfrak{C}_{21}^3(\tau_k) = \underset{0\leqq\lambda\leqq 2\pi}{\mathfrak{S}} E[w_k = w_k(\tau_1e^{i\lambda}, z_2),\ |z_2|\leqq\tau_2],$$

$$(7.3.2)\qquad \mathfrak{C}_{22}^3(\tau_k) = \underset{0\leqq\lambda\leqq 2\pi}{\mathfrak{S}} E[w_k = w_k(z_1, \tau_2e^{i\lambda}),\ |z_1|\leqq\tau_1],$$

and their intersection $\mathfrak{F}_2^2(\tau_k)=\mathfrak{C}_{21}^3(\tau_k)\cdot\mathfrak{C}_{22}^3(\tau_k)$ is the distinguished boundary surface of $\mathfrak{B}_2(\tau_k)$.

We shall find it convenient to define a real (uniformly) continuous function $G_2(w_1, w_2; \tau_k)$ in $\mathfrak{B}_2(P_k)$ which is positive in $\mathfrak{B}_2(\tau_k)$, nonpositive in $\mathfrak{B}_2(P_k)-\mathfrak{B}_2(\tau_k)$, and, therefore, zero on the boundary of $\mathfrak{B}_2(\tau_k)$. The existence of such functions follows from the fact that $\mathfrak{B}_2(\tau_k)$ is an image of a bicylinder $\mathfrak{B}_1(\tau_k)$; we can define an analogous function $G_1(z_1, z_2; \tau_k)$ in $\mathfrak{B}_1(P_k)$, and then $G_1[z_1(w_1, w_2), z_2(w_1, w_2); \tau_k]$ is a function possessing the indicated properties. However, given $\mathfrak{B}_2(\tau_k)$, $G_2(w_1, w_2; \tau_k)$ can be constructed without using the PT $\mathbf{w}^{-1}$.

7.4. We now discuss the surface

$$(7.4.1)\qquad \mathfrak{S}_2^2(\psi_1, \psi_2) = E[\arg w_k = \psi_k = \text{const.},\ (w_1, w_2)\in\mathfrak{B}_2(P_k)].$$

---

(20) The double subscripts are written without a comma; for example $\mathfrak{b}_{21}^3$ for $\mathfrak{b}_{2,1}^3$, the *first*, number, 1 or 2, indicating the space (z_1z_2) or (w_1w_2). See §2.

LEMMA 2. *There exists a set*(21) $\mathfrak{H}^2$, $m(\mathfrak{H}^2)=4\pi^2$, *lying in the square* $E[0\leqq\psi_k<2\pi]$ *such that, if* $(\psi_1, \psi_2)\in\mathfrak{H}^2$, $\mathfrak{B}_2^2\cdot\mathfrak{S}_2^2(\psi_1, \psi_2)$ *is composed of only finitely many points.*

We remark that, since the number of branch points of the second kind is finite, we can further suppose that $\mathfrak{S}_2^2(\psi_1, \psi_2)$ contains no such points if $(\psi_1, \psi_2)\in\mathfrak{H}^2$.

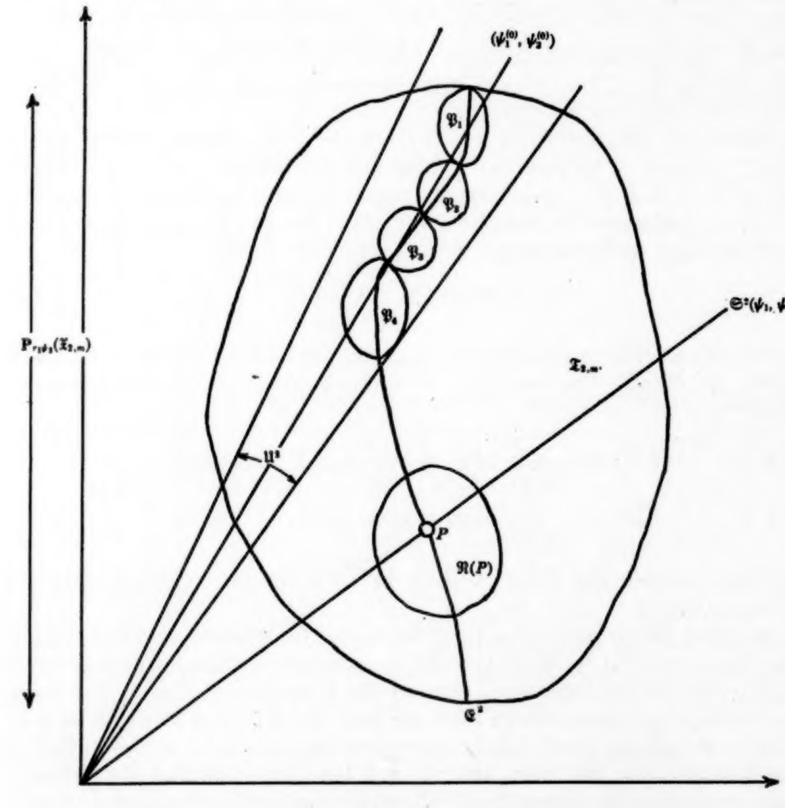


FIG. 2

Denoting branch points of the second kind by $(w_1^{(\nu)}, w_2^{(\nu)})$, $\nu=1, 2, \cdots, p$, we can (by the Weierstrass preparation theorem and the Heine-Borel theorem) cover the domain $\mathfrak{B}_2(P_k)-\mathfrak{S}_{\nu=1}^{p}E[|w_1-w_1^{(\nu)}|^2+|w_2-w_2^{(\nu)}|^2<\epsilon]$ with a

(21) $m(\mathfrak{H}^2)$ denotes the measure of $\mathfrak{H}^2$.

finite number of domains $\mathfrak{T}_{2,m}$ in such a way that in each $\bar{\mathfrak{T}}_{2,m}$ (the closure of $\mathfrak{T}_{2,m}$) the branch surface $\mathfrak{E}_2^2$ can be represented in the form([22])

$$w_1 = H(w_2), \tag{7.4.2}$$

or

$$w_2 = H(w_1),$$

where H is an analytic function of one complex variable regular in the projection $\mathbf{P}_{w_k}[\bar{\mathfrak{T}}_{2,m}]$ of $\bar{\mathfrak{T}}_{2,m}$ on the w_k-plane (see §2).

Suppose that the first representation of (7.4.2) is valid for $\mathfrak{E}_2^2 \cdot \mathfrak{T}_{2,m}$. Then (7.4.2) can be written in the form

$$r_1 = [H(r_2 e^{i\psi_2})\bar{H}(r_2 e^{-i\psi_2})]^{1/2}, \quad \psi_1 = (1/2i)[\lg H(r_2 e^{i\psi_2}) - \lg \bar{H}(r_2 e^{-i\psi_2})]. \tag{7.4.3}$$

Since $H \neq 0$, r_1 and ψ_1 are one-valued and regular functions of r_2 and ψ_2. In the neighborhood $\mathfrak{N}(P)$ of any point P of $\mathfrak{E}_2^2 \cdot \mathfrak{T}_{2,m}$ in which

$$D(r_2, \psi_2) = \partial[\lg H(r_2 e^{i\psi_2}) - \lg \bar{H}(r_2 e^{-i\psi_2})]/\partial r_2 \neq 0, \tag{7.4.4}$$

the surface $\mathfrak{E}_2^2 \cdot \mathfrak{T}_{2,m}$ can also be written in the form

$$r_k = t_k(\psi_1, \psi_2), \tag{7.4.5}$$

where the t_k are one-valued functions of ψ_1, ψ_2, so that $\mathfrak{E}_2^2 \cdot \mathfrak{S}_2^2(\psi_1, \psi_2) \cdot \mathfrak{N}(P)$ consists of only *one* point.

We shall determine for what values of (ψ_1, ψ_2) the function $D(r_2, \psi_2)$ vanishes on $\mathfrak{E}_2^2 \cdot \mathfrak{T}_{2,m}$. Since $D(r_2, \psi_2)$ is an analytic function of 2 real variables r_2, ψ_2, regular in $\mathbf{P}_{r_2\psi_2}(\bar{\mathfrak{T}}_{2,m})$, we can cover $\mathbf{P}_{r_2\psi_2}(\bar{\mathfrak{T}}_{2,m})$ with a finite number of neighborhoods $\mathfrak{U}_{m,\mu}^2$ in each of which the equation $D(r_2, \psi_2) = 0$ can be expressed in the form

$$\overset{(m\mu)}{\chi_0}(\psi_2) r_2^{p} + \overset{(m\mu)}{\chi_1}(\psi_2) r_2^{p-1} + \cdots + \overset{(m\mu)}{\chi_p}(\psi_2) = 0. \tag{7.4.6}$$

Since $\chi_0^{(m\mu)}(\psi_2)$ has in every $\mathfrak{U}_{m,\mu}^2$ only a finite number of factors $(\psi_2 - \psi_2^{(m\mu l)})$, we can cover $\mathbf{P}_{r_2\psi_2}(\bar{\mathfrak{T}}_{2,m})$ (perhaps multiply) with a finite number of domains $\mathfrak{M}_{m\mu s}^2$ such that $D(r_2, \psi_2) = 0$ can be written in $\mathfrak{M}_{m\mu s}^2$ in the form

$$r_2 = \overset{(m\mu s)}{r_2}(\psi_2) \qquad (\psi_2 \neq \overset{(m\mu l)}{\psi_2}), \tag{7.4.7}$$

where the $r_2^{(m\mu s)}(\psi_2)$ are algebroid functions in $\mathbf{P}_{\psi_2}(\mathfrak{M}_{m\mu s}^2)$. Substituting from (7.4.7) into the second equation of (7.4.3), we see that $D(r_2, \psi_2)$ vanishes on $\mathfrak{E}_2^2 \cdot \bar{\mathfrak{T}}_{2,m}$ only for values of (ψ_1, ψ_2) for which

$$\psi_1 = A^{(m\mu s)}(\psi_2), \tag{7.4.8}$$

or

([22]) See Osgood [1, p. 113]. Figure 2 is a schematic diagram in which four-dimensional domains have been replaced by plane domains, and surfaces by curves.

$$\psi_2 = \psi_2^{(m\mu l)},$$

where the $A^{(m\mu s)}(\psi_2)$ are again algebroid functions in $\mathbf{P}_{\psi_2}(\mathfrak{M}^2_{m\mu s})$.

In every $\mathbf{P}_{\psi_1\psi_2}(\mathfrak{T}_{2,m})$ each expression (7.4.8) defines a curve the two-dimensional measure of which is zero. Since the number of these curves is finite and since m, μ, s, l run through a finite set of integers, the measure of the set $\mathfrak{H}^2$ which is left after removing all these curves from $\mathrm{E}[0 \leqq \psi_k < 2\pi]$ will be $4\pi^2$.

Consider now a neighborhood $\mathfrak{U}^2_0$ of a point $(\psi_1^{(0)}, \psi_2^{(0)}) \in \mathfrak{H}^2$. We can choose $\mathfrak{U}^2_0$ so small that $\mathfrak{Y}^2_0 \subset \mathfrak{H}^2$ (where $\mathfrak{Y}^2_0$ is the closure of $\mathfrak{U}^2_0$). Since in every domain $\mathfrak{K}_m = \mathfrak{T}_{2,m} \cdot \mathrm{E}[(\psi_1, \psi_2) \in \mathfrak{Y}^2_0]$, we have $D(r_2, \psi_2) \neq 0$ on $\mathfrak{K}_m \cdot \mathfrak{E}^2_2$, we can cover $\mathfrak{K}_m$ with a finite number of neighborhoods $\mathfrak{P}_\mu$ in each of which we have the representation (7.4.5) for $\mathfrak{P}_\mu \cdot \mathfrak{E}^2_2 \cdot \mathfrak{S}^2_2(\psi_1, \psi_2)$. Since the $t_k(\psi_1, \psi_2)$ are one-valued functions, and since there is only a finite number of m and μ, $\mathfrak{E}^2_2 \cdot \mathfrak{S}^2_2(\psi_1, \psi_2)$ consists of a finite number of points for $(\psi_1, \psi_2) \in \mathfrak{U}^2_0$, and this proves Lemma 2.

7.5. In this section we study the curves (for $\mathfrak{E}^3_{21}(\tau_1)$ see (7.3.1))

$$\overset{1}{\mathfrak{c}}_{21}(\tau_1) = \overset{2}{\mathfrak{S}}_2(\psi_1, \psi_2) \cdot \overset{3}{\mathfrak{E}}_{21}(\tau_1), \tag{7.5.1}$$

on which we suppose no branch points (that is, points of $\mathfrak{B}^2_2$) are situated, and where (ψ_1, ψ_2) is a point of $\mathfrak{H}^2$. By (7.3.1) we may write (7.5.1) in the form

$$\overset{1}{\mathfrak{c}}_{21}(\tau_1) = \mathrm{E}[\tau_1 = | z_1(r_1 e^{i\psi_1}, r_2 e^{i\psi_2}) |]. \tag{7.5.2}$$

If $r_k > 0$, we call a point (r_1, r_2) of the curve $\overset{1}{\mathfrak{c}}_{21}(\tau_1)$ for which (simultaneously)

$$\partial(| z_1 |)/\partial r_1 = 0, \qquad \partial(| z_1 |)/\partial r_2 = 0, \tag{7.5.3}$$

a "node." A node is either an isolated point or a point at which (finitely) many different branches of the curve meet. If (r_1^0, r_2^0) is neither a node nor a point of $\mathfrak{B}^2_2$, then, by a well known theorem of implicit function-theory, we can represent $\overset{1}{\mathfrak{c}}_{21}(\tau_1)$ in a neighborhood of (r_1^0, r_2^0) in one of the two following forms:

$$r_1 = a(r_2), \qquad r_2 = a(r_1), \tag{7.5.4}$$

where a is in both cases an analytic function of one real variable.

It follows, in particular, that if the curve $\overset{1}{\mathfrak{c}}_{21}(\tau_1)$, $\overset{1}{\mathfrak{c}}_{21}(\tau_1) \cdot \mathfrak{B}^2_2 = 0$, ends in the interior of $\mathfrak{S}^2_2(\psi_1, \psi_2)$, it must end at a node. A piece of $\overset{1}{\mathfrak{c}}_{21}(\tau_1)$ which contains no node but which connects two nodes, two boundary points of $\mathfrak{S}^2_2(\psi_1, \psi_2)$, or a boundary point and a node, we call a branch of $\overset{1}{\mathfrak{c}}_{21}(\tau_1)$. Each branch is a regular curve([23]).

Since the number of nodes is finite, and since finitely many branches meet

([23]) That is to say, a curve whose equation is differentiable infinitely often. In particular, a branch has a tangent at every point including the end points (since at a node we have the representation (7.5.4), where $a(r_k)$ is an algebroid function).

in a node, we see that there are only a *finite number of branches of which at least one end point is a node.*

The neighborhood on $\mathfrak{S}_2^2(\psi_1, \psi_2)$ of any interior point of a branch may consist of three kinds of points: (i) points corresponding to $|z_1| = \tau_1$, that is, points of $\mathfrak{c}_{21}^1(\tau_1)$; (ii) points corresponding to $|z_1| < \tau_1$; and (iii) points corresponding to $|z_1| > \tau_1$. We call points of the second kind $L(\tau_1)$-points, points of the third kind $M(\tau_1)$-points. We then say that a branch is of the first kind if L-points[24] lie on one side of it, M-points on the other; of the second (third) kind if L-points (M-points) lie on both sides of it.

7.6. By means of the function $G_2(w_1, w_2; 1)$ introduced above (§7.3), we divide all points of $\mathfrak{S}_2^2(\psi_1, \psi_2)$ into three categories; namely, i-points, e-points and a-points, at which $G_2(w_1, w_2; 1)$ is positive, zero, or negative, respectively.

If $(\psi_1, \psi_2) \in \mathfrak{H}^2$ (see §7.4), the totality of i-points of $\mathfrak{S}_2^2(\psi_1, \psi_2)$ forms a set which lies on a Riemann surface. By Lemma 2 there are at most finitely many branch points in this set, and each of them may be uniformized by (7.1.3). This set will consist in general of many disconnected components. Since (see §6) the mapping is schlicht in

$$\mathfrak{D}_1 \cdot [|z_1| \leq \epsilon] + \mathfrak{D}_1 \cdot [|z_2| \leq \epsilon],$$

and is the identical transformation $w_k = z_k$ on $|z_1| = 0$ and $|z_2| = 0$, there is one and only one component which contains the lines

$$\overset{\dagger 1}{\mathfrak{a}_{21}} = E[0 \leq r_1 \leq 1, r_2 = 0], \overset{\dagger 1}{\mathfrak{a}_{22}} = E[r_1 = 0, 0 \leq r_2 \leq 1]$$

as part of its boundary. We denote this component by $\mathfrak{S}_2^{\dagger 2}(\psi_1, \psi_2)$. The boundary of $\mathfrak{S}_2^{\dagger 2}(\psi_1, \psi_2)$ consists of $\mathfrak{a}_{21}^{\dagger 1}$ and $\mathfrak{a}_{22}^{\dagger 1}$, and of pieces of $\mathfrak{c}_{21}^1(1)$ and $\mathfrak{c}_{22}^1(1)$ which we shall denote by $\mathfrak{c}_{21}^{\dagger 1}$, $\mathfrak{c}_{22}^{\dagger 1}$, respectively (see Figure 3). We shall require the following lemma.

LEMMA 3. *If $(\psi_1, \psi_2) \in \mathfrak{H}^2$, the projection of the surface*

$$\overset{\dagger 2}{\mathfrak{S}_1}(\psi_1, \psi_2) = \mathbf{w}^{-1}[\overset{\dagger 2}{\mathfrak{S}_2}(\psi_1, \psi_2)] \subset E[\arg w_k(\rho_1 e^{i\theta_1}, \rho_2 e^{i\theta_2}) = \psi_k = \text{const.}]$$

on the $\rho_1\rho_2$-plane [i.e., the totality of the (ρ_1, ρ_2) coordinates of points of $\mathfrak{S}_1^{\dagger 2}(\psi_1, \psi_2)$] covers a set

$$\mathfrak{J}^2 \subset E[0 < \rho_\kappa < 1],$$

where $\mathrm{m}(\mathfrak{J}^2) = 1$.

In fact, suppose that $0 < \lambda < 1$, and that there are no points of $\mathfrak{B}_2^2$ on $\mathfrak{c}_{21}^1(\lambda)$. Since the mapping is schlicht in a neighborhood of the line $|z_1| = \lambda$, $|z_2| = 0$, one and only one branch $\mathfrak{a}_1^1$ of $\mathfrak{c}_{21}^1(\lambda)$ begins at the point $K_0 = [r_1 = \lambda, r_2 = 0]$. We shall show that by adding to $\mathfrak{a}_1^1$ further suitably chosen branches of $\mathfrak{c}_{21}^1(\lambda)$, we obtain a connected curve which cuts $\mathfrak{c}_{22}^1(1)$.

(24) We shall omit the parameter τ_1 in $L(\tau_1)$ and $M(\tau_1)$ hereafter.

Beginning at K_0 we travel outward along $\mathfrak{a}_1^1$ keeping L-points to the left until we reach a point of $\mathfrak{a}_1^1$ corresponding to

$$[|z_1(w_1, w_2)| = \lambda, \ |z_2(w_1, w_2)| = 1],$$

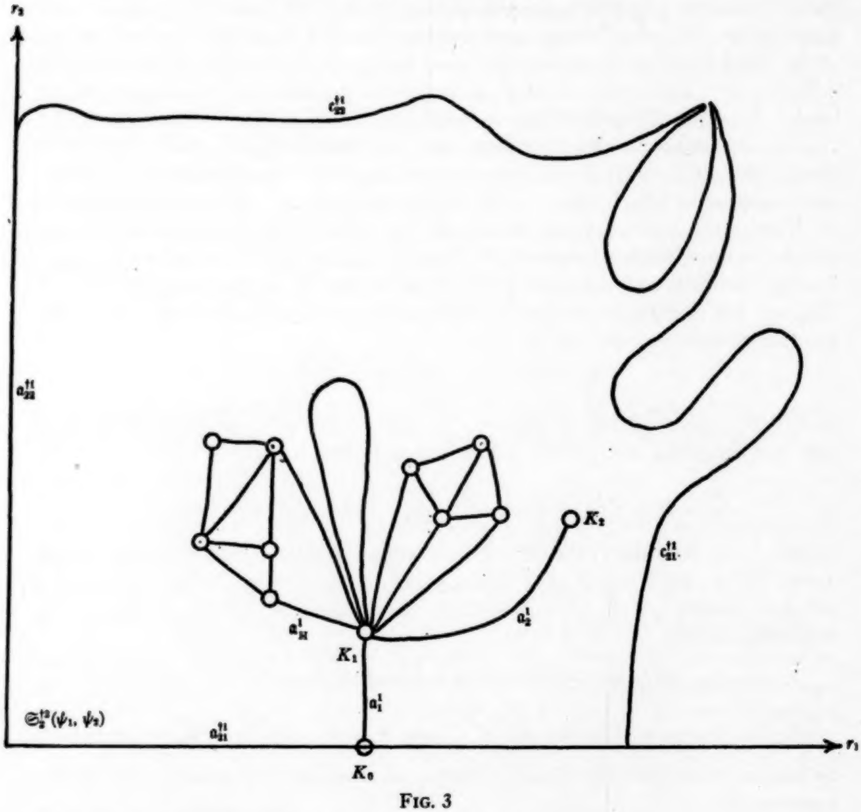


FIG. 3

or a node K_1 for which(25)

$$|z_1| = \lambda, \ |z_2| < 1.$$

In the latter case we start from the L-edge of $\mathfrak{a}_1^1$ and turn clockwise round a sufficiently small circle(26) with center K_1 until we meet a new branch $\mathfrak{a}_H^1$ of the first kind. $\mathfrak{a}_H^1$ exists; otherwise we should arrive at the end of $\mathfrak{a}_1^1$ where

(25) Hereafter, for simplicity, we shall write $|z_1| = \lambda$, $|z_2| = 1$, and omit the argument (w_1, w_2).

(26) The radius of the circle depends, of course, on the node. It is small enough so that it intersects branches with one end point at K_1 once and only once, branches with two end points at K_1 only twice, and intersects no branch which does not have an end point at K_1.

M-points are situated. This is impossible since we cannot connect a point L with a point M without meeting a branch of the first kind([27]).

We now travel along $\mathfrak{a}^1_{\mathrm{H}}$ with L-points to the left until we meet a point corresponding to $[|z_1| = \lambda, |z_2| = 1]$, or a new node at which $|z_1| = \lambda$, $|z_2| < 1$, and repeat the same process. It may happen that after running over a set of branches $(\mathfrak{a}^1_{\mathrm{H}}, \mathfrak{a}^1_{\mathrm{H}+1}, \cdots, \mathfrak{a}^1_{\mathrm{H}+\nu})$, $\nu \geqq 0$, we arrive back at the node K_1 from which we started without leaving the domain $\mathrm{E}[|z_1| < 1, |z_2| < 1]$. If a set of branches chosen in this way has the property that only the first and the

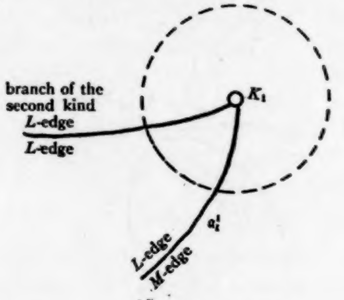


FIG. 4

last have end points at K_1([28]), we call it a chain $\mathfrak{u}^1_1(K_1)$ hung from K_1. Since in turning clockwise round K_1 from $\mathfrak{a}^1_1$ to $\mathfrak{a}^1_{\mathrm{H}}$ we meet no branch of the first kind, we conclude that $a^1_{\mathrm{H}+\nu}$ must lie to the right of $\mathfrak{a}^1_{\mathrm{H}}$. The chain $\mathfrak{u}^1_1(K_1)$ has an important property: It is possible to draw a curve from the L-side of $\mathfrak{u}^1_1(K_1)$ which begins and ends at K_1, but which lies (apart from its end points and perhaps points lying on branches of the second kind) completely in an L-domain, and which, therefore, cuts $|z_1| = \lambda$ (if at all) only in branches of the second kind. This follows from the fact that in forming the chain we turn round each node K_p in the clockwise sense until we meet a branch of the first kind.

After traveling over the branch $\mathfrak{a}^1_{\mathrm{H}+\nu}$ which leads to K_1, we move clockwise round a sufficiently small circle until we again meet a branch $\mathfrak{a}^1_i$, say, of the first kind, then continue along $\mathfrak{a}^1_i$, and so on. But we may now run over a second chain $\mathfrak{u}^1_2(K_1)$ which leads us over finitely many branches (situated in $\mathrm{E}[|z_1| < 1, |z_2| < 1])$ and then back to K_1 again. *However, the chains* $\mathfrak{u}^1_1(K_1)$ *and* $\mathfrak{u}^1_2(K_1)$ *have no common node except* K_1. For, in the neighborhood of K_1, $\mathfrak{a}^1_i$ lies on the L-side of $\mathfrak{u}^1_1(K_1)$. If, therefore, we move from K_1 along $\mathfrak{a}^1_i$, we

---

([27]) For we must meet a branch. If the branch were of the second or the third kind we could connect an L- with an M-point without passing through a point for which $|z_1| = \lambda$.

([28]) If the chain contains but one branch $\mathfrak{a}^1_{\mathrm{H}}$, we make a distinction between its two ends.

must meet a branch $\mathfrak{a}^1_{i+\nu}$ which lies on the L-side of $\mathfrak{u}^1_1(K_1)$ and ends in a node K_n at which two branches of the first chain $\mathfrak{a}^1_{H+\mu}$ and $\mathfrak{a}^1_{H+\mu+1}$ also end (see Figure 5). This is impossible; otherwise in turning clockwise round K_n, after having travelled $\mathfrak{a}^1_{H+\mu}$, we should have met $\mathfrak{a}^1_{i+\nu}$, and *not* $\mathfrak{a}^1_{H+\mu+1}$.

Continuing the process, we may travel over a set $\mathcal{C}(K_1)$ of chains $\mathfrak{u}^1_1(K_1)$, $\mathfrak{u}^1_2(K_1), \cdots$, hung from K_1; but no two chains of $\mathcal{C}$ have a common node other than K_1. Since in turning from $\mathfrak{a}^1_1$ clockwise round K_1 through 360° we meet the initial before we meet the final branch of a chain, and since the number of branches attached to a node is finite, we see that the number

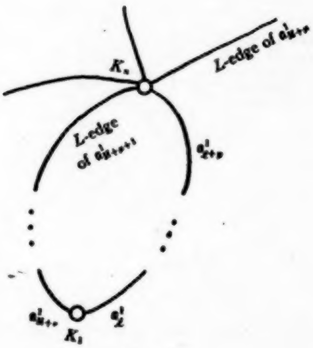


FIG. 5

of chains in $\mathcal{C}(K_1)$ is finite. By traveling round all chains hung from K_1 we must therefore find a branch, $\mathfrak{a}^1_2$, say, which will take us away from K_1 never to return; otherwise we could connect the L-edge of the last chain with the M-edge of $\mathfrak{a}^1_1$ without passing through a branch of the first kind, and this is impossible. $\mathfrak{a}^1_2$ leads us to a node K_2. At K_2 we proceed in the same manner as at K_1 with $\mathfrak{a}^1_1$ replaced by $\mathfrak{a}^1_2$. We cannot travel along $\mathfrak{a}^1_2$ a second time without passing through K_1 again, and this is again impossible. The continuation of our process, therefore, always leads us to new nodes, and since the number of nodes is finite, we must reach a branch on which a point corresponding to $|z_1| = \lambda$, $|z_2| = 1$ lies. If not we should pass through all nodes and reach a branch $\mathfrak{a}^1_s$ of which one end point is a boundary point of $\mathfrak{S}^2_2(\psi_1, \psi_2)$. This end point corresponds neither to $|z_1| = 0$ nor to $|z_1| = \mathrm{P}_1$; for our curve corresponds to $|z_1| = \lambda$, $0 < \lambda < 1 < \mathrm{P}_1$. Furthermore, it cannot be a point for which $|z_2| = 0$; for we started from the point $K_0(r_1 = \lambda, r_2 = 0)$ where, by hypothesis (see p. 147) there is only one branch, and we could, therefore, arrive at K_0 only by passing through K_1. It follows that the branch $\mathfrak{a}^1_s$ has one end at a point on $|z_2| = \mathrm{P}_2$, or at a node for which $|z_2| > 1$.

The system of branches[29]

$$\mathfrak{l}_2^1(\lambda) = (\mathfrak{a}_1^1, \mathfrak{a}_2^1, \cdots, \mathfrak{a}_{s-1}^1, \mathfrak{a}_{s,0}^1)$$

together forms a connected curve composed only of i-points.

Since, if ϵ is small enough, we can connect $\mathfrak{l}_2^1(\lambda)$ to a neighborhood of the point $r_1=0$, $r_2=0$, by means of the curve

$$E[|w_2| = \epsilon, \epsilon \leqq |w_1| \leqq \lambda]$$

we see that $\mathfrak{l}_2^1(\lambda)$ lies in $\mathfrak{S}_2^{\dagger 2}(\psi_1, \psi_2)$. The curve $\mathfrak{l}_1^1(\lambda) = \mathbf{w}^{-1}[\mathfrak{l}_2^1(\lambda)]$ is by Lemma 1 a connected curve with end points at $[|z_1| = \lambda, |z_2| = 0]$ and $[|z_1| = \lambda, |z_2| = 1]$, and its projection on the $\rho_1\rho_2$-plane is therefore *a fortiori* a connected curve with end points at $[\rho_1 = \lambda, \rho_2 = 0]$, $[\rho_1 = \lambda, \rho_2 = 1]$. Since $\rho_1 = \lambda$ on the curve (by definition), the projection of $\mathfrak{l}_1^1(\lambda)$ is a straight line $E[\rho_1 = \lambda, 0 < \rho_2 < 1]$.

Since, by Lemma 2, $\mathfrak{S}_2^2(\psi_1, \psi_2)$ contains only a finite number of points $(w_1^{(H)}, w_2^{(H)})$ of $\mathfrak{B}_2^2$, there exists only a finite number of $\lambda^{(H)} = |z_1(w_1^{(H)}, w_2^{(H)})|$ for which

$$\mathfrak{c}_2^1(\lambda^{(H)}) = E[\lambda^{(H)} = |z_1(w_1, w_2)|, \arg w_k = \psi_k]$$

contains a point $(w_1^{(H)}, w_2^{(H)})$ of $\mathfrak{B}_2^2$. For any other value of λ, and hence for almost all λ, the above argument is valid. Since for any λ except $\lambda = \lambda^{(H)}$, $\mathfrak{l}_1^1(\lambda)$ is contained in $\mathfrak{S}_1^{\dagger 2}(\psi_1, \psi_2)$, the projection of $\mathfrak{S}_1^{\dagger 2}(\psi_1, \psi_2)$ covers $E(0 < \rho_k < 1)$ with exception of a null set. This completes the proof of Lemma 3.

7.7. In §4 we defined $\mathfrak{S}_2^{\dagger 2}(\psi_1, \psi_2)$, $(\psi_1, \psi_2) \in \mathfrak{H}^2$, as that connected component of the set of i-points of $\mathfrak{S}_2^2(\psi_1, \psi_2)$ the boundary of which contains the point $\mathbf{w}(0,0)$. We now define

$$\mathfrak{B}_2^{\dagger} = \mathop{\mathfrak{S}}_{(\psi_1, \psi_2) \in \mathfrak{H}^2} \mathfrak{S}_2^{\dagger 2}(\psi_1, \psi_2). \tag{7.7.1}$$

We remark, however, that our theorem is true for any domain

$$\mathfrak{B}_2^{\dagger\dagger} = \mathop{\mathfrak{S}}_{(\psi_1, \psi_2) \in \mathfrak{H}^2} \mathfrak{S}_2^{\dagger\dagger 2}(\psi_1, \psi_2) \tag{7.7.2}$$

which has the property that the projection of the surface

$$\mathfrak{S}_1^{\dagger\dagger 2}(\psi_1, \psi_2) = \mathbf{w}^{-1}[\mathfrak{S}_2^{\dagger\dagger 2}(\psi_1, \psi_2)] \tag{7.7.3}$$

fills the unit square $E[0 < \rho_k < 1]$.

For example, from §7.6 we see that we may take (in (7.7.2)) the surface

$$\mathfrak{S}_2^{\dagger\dagger 2}(\psi_2, \psi_2) = \mathop{\mathfrak{S}}_{0 < \lambda < 1} \mathfrak{l}_2^1(\lambda). \tag{7.7.4}$$

[29] $\mathfrak{a}_{s,0}^1$ here denotes only the portion of the last branch for which $|z_2| < 1$.

However, the resulting domain $\mathfrak{B}^{\dagger\dagger}$ has the disadvantage that it is defined by means of the PT $\mathbf{w}$, whereas the domain $\mathfrak{B}^{\dagger}$ is obtained from $\mathfrak{B}_2$ by a geometrical operation definable without reference to $\mathbf{w}$.

8. Proof. Analytical part

8.1. We begin by restating the definition of B-area (see Bergman [2]). Let

$$(8.1.1)\qquad \begin{aligned}\mathfrak{L}^2 = \mathrm{E}[Z_1(u_1, u_2) &= \chi_1(u_1, u_2) + i\chi_2(u_1, u_2),\\ Z_2(u_1, u_2) &= \chi_3(u_1, u_2) + i\chi_4(u_1, u_2),\ 0 \leqq u_k \leqq 1],\end{aligned}$$

the $\chi_j(u_1, u_2)$ being real functions of two real variables which have continuous derivatives. Then $\mathrm{B}(\mathfrak{L}^2)$, the B-area of $\mathfrak{L}^2$, is defined by

$$(8.1.2)\qquad \mathrm{B}(\mathfrak{L}^2) = \int_0^1 \int_0^1 \mathrm{b}(Z_1, Z_2)du_1du_2, \quad \mathrm{b}(Z_1, Z_2) = |\,\partial(Z_1, Z_2)/\partial(u_1, u_2)\,|\,;$$

whereas $\mathrm{A}(\mathfrak{L}^2)$, the ordinary area of it, is defined by

$$(8.1.3)\qquad \int_0^1 \int_0^1 \mathrm{a}(Z_1, Z_2)du_1du_2,$$

where[30]

$$\mathrm{a}(Z_1, Z_2) = EG - F^2 = \begin{vmatrix} \sum \chi_{ju_1}^2 & \sum \chi_{ju_1}\chi_{ju_2} \\ \sum \chi_{ju_1}\chi_{ju_2} & \sum \chi_{ju_2}^2 \end{vmatrix}, \quad \sum = \sum_{j=1}^{4}, \quad \chi_{ju_k} = \frac{\partial \chi_j}{\partial u_k}.$$

We have, therefore,

$$[\mathrm{b}(Z_1, Z_2)]^2 = EG - F^2 - T^2,$$

where

$$(8.1.4)\qquad \begin{aligned} T &= -\chi_{1u_1}\chi_{2u_2} + \chi_{2u_1}\chi_{1u_2} - \chi_{3u_1}\chi_{4u_2} + \chi_{4u_1}\chi_{3u_2} \\ &= \frac{1}{4i}\sum_{k=1}^{2}\left[-\left(\frac{\partial Z_k}{\partial u_1} + \frac{\partial \overline{Z}_k}{\partial u_1}\right)\left(\frac{\partial Z_k}{\partial u_2} - \frac{\partial \overline{Z}_k}{\partial u_2}\right)\right. \\ &\qquad \left. + \left(\frac{\partial Z_k}{\partial u_2} + \frac{\partial \overline{Z}_k}{\partial u_2}\right)\left(\frac{\partial Z_k}{\partial u_1} - \frac{\partial \overline{Z}_k}{\partial u_1}\right)\right] \\ &= \mathrm{Im}\left[\sum_{k=1}^{2} \frac{\partial Z_k}{\partial u_1}\frac{\partial \overline{Z}_k}{\partial u_2}\right], \end{aligned}$$

and hence[31]

---

(30) E, F and G are the coefficients of the fundamental form for the line-element.

(31) See Bergman [2, p. 476].

$$\begin{aligned}\mathrm{b}(Z_1, Z_2) &= (EG - F^2 - T^2)^{1/2} \geqq \mathrm{a}(Z_1, Z_2) - |T| \\ &\geqq \mathrm{a}(Z_1, Z_2) - \sum_{k=1}^{2}\left|\frac{\partial Z_k}{\partial u_1}\frac{\partial Z_k}{\partial u_2}\right|.\end{aligned} \tag{8.1.5}$$

8.2. We recall that $w_k = r_k e^{i\psi_k}$; write $\mathfrak{S}_{2,\lambda}^{\dagger 2}(\psi_1, \psi_2)$ for the portion of the surface $\mathfrak{S}_2^{\dagger 2}(\psi_1, \psi_2)$ for which $2|w_1 w_2| > \lambda$, and write $\mathbf{L}_1$, $\mathbf{L}_2$ for the two PT's $Z_k = \lg z_k$, $W_k = \lg w_k$, respectively. We then define

$$\begin{aligned}B_{1,\lambda}(\psi_1, \psi_2) &= \mathrm{B}\{\mathbf{L}_1\mathbf{w}^{-1}[\mathfrak{S}_{2,\lambda}^{\dagger 2}(\psi_1, \psi_2)]\} \\ &= \iint_{\mathfrak{S}_{2,\lambda}^{\dagger 2}(\psi_1,\psi_2)} \left|\frac{\partial(Z_1, Z_2)}{\partial(\lg w_1, \lg w_2)}\right| d(\lg w_1) d(\lg w_2) \\ &= \iint_{\mathfrak{S}_{2,\lambda}^{\dagger 2}(\psi_1,\psi_2)} \left|\frac{\partial(Z_1, Z_2)}{\partial(r_1, r_2)}\right| dr_1 dr_2.\end{aligned} \tag{8.2.1}$$

8.3. We prove the following lemma.

LEMMA 4. *Suppose that hypothesis* (5.1.2) *of the theorem is satisfied. Then, if* $0 < \lambda < 2$,

$$\begin{aligned}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B_{1,\lambda}(\psi_1, \psi_2) d\psi_1 d\psi_2 \geqq 2\pi^2 \lg^2 \lambda + 4\pi^2 \lg 2 \cdot \lg (1/\lambda) - \alpha_2 \\ - O[\lambda \lg (1/\lambda)].\end{aligned} \tag{8.3.1}$$

Taking $u_k = r_k$, we have from (8.1.5) and (8.2.1) that

$$\begin{aligned}B_{1,\lambda}(\psi_1, \psi_2) \geqq \iint_{\mathfrak{S}_{2,\lambda}^{\dagger 2}(\psi_1,\psi_2)} \mathrm{a}(Z_1, Z_2) dr_1 dr_2 \\ - \iint_{\mathfrak{S}_2^{\dagger 2}(\psi_1,\psi_2)} \mathrm{e}(w_1, w_2) dr_1 dr_2,\end{aligned} \tag{8.3.2}$$

where $Z_k = \lg [z_k(r_1 e^{i\psi_1}, r_2 e^{i\psi_2})]$, and

$$\mathrm{e}(w_1, w_2) = \sum_{k=1}^{2} |z_k|^{-2} |(\partial z_k/\partial w_1) \cdot (\partial z_k/\partial w_2)|.$$

Let

$$\mathfrak{P}_{1,\lambda}^{2} = \mathbf{P}_{\lg \rho_1, \lg \rho_2}\{\mathbf{L}_1\mathbf{w}^{-1}[\mathfrak{S}_{2,\lambda}^{\dagger 2}(\psi_1, \psi_2)]\}, \tag{8.3.3}$$

$$\mathfrak{T}_{1,\lambda}^{2} = \mathrm{E}[\lambda \leqq 2\rho_1\rho_2 < 2, \ \rho_k < 1]. \tag{8.3.4}$$

Since by (6.2.3)

$$2|z_1(w_1, w_2) \cdot z_2(w_1, w_2)| \leqq \lambda + O(\lambda^2)$$

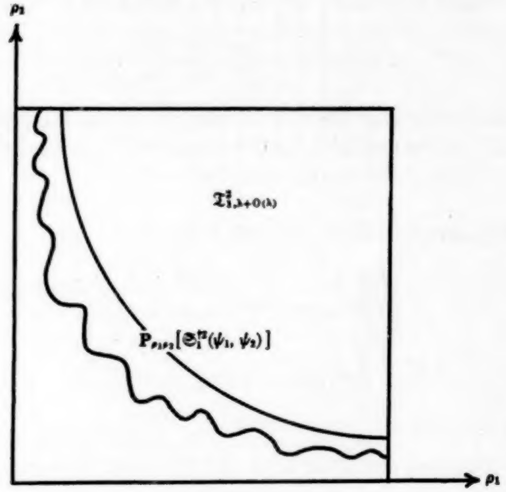


FIG. 6

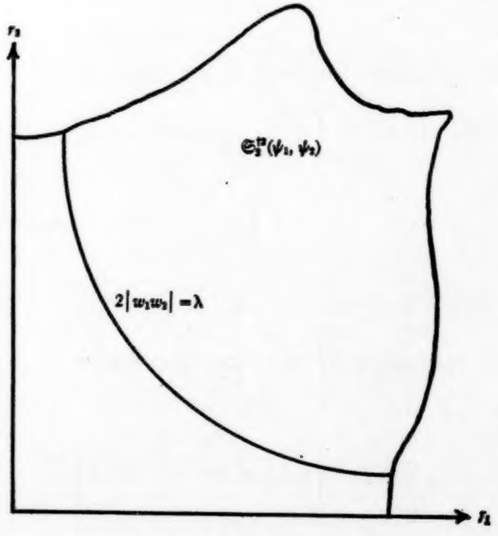


FIG. 7

for $2|w_1w_2| = \lambda$ we see that the boundary of $\mathfrak{P}^2_{1,\lambda}$ lies *outside* $\mathbf{L}_1[\mathfrak{T}^2_{1,\lambda+O(\lambda^3)}]$

(see Figure 8). Suppose now that $(\psi_1, \psi_2) \in \mathfrak{H}^2$, $m(\mathfrak{H}^2) = 4\pi^2$, where $\mathfrak{H}^2$ is the set in Lemma 3. Then by Lemma 3 every point of $L_1[\mathfrak{T}^2_{1,\lambda+O(\lambda^2)}]$, with exception of a set of measure zero, is a point of $\mathfrak{P}^2_{1,\lambda}$; and so

$$\begin{aligned} A[\mathfrak{P}^2_{1,\lambda}] &\geqq A[L_1(\mathfrak{T}^2_{1,\lambda+O(\lambda^2)})] = 1/2[\lg (\lambda + O(\lambda^2)) - \lg 2]^2 \\ &\geqq (1/2) \lg^2 \lambda + \lg 2 \cdot \lg (1/\lambda) - O[\lambda \lg (1/\lambda)]. \end{aligned} \tag{8.3.5}$$

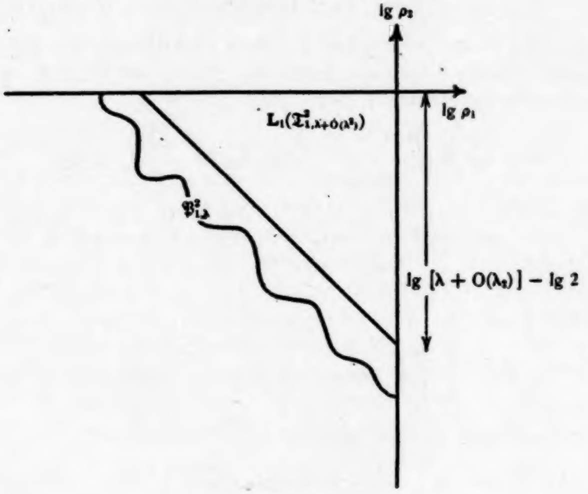


FIG. 8

Hence by (8.3.5) we have for $(\psi_1, \psi_2) \in \mathfrak{H}^2$, and therefore for almost all (ψ_1, ψ_2), that

$$\begin{aligned} \iint_{\mathfrak{S}^{\dagger 2}_{2,\lambda}(\psi_1,\psi_2)} a(Z_1, Z_2) dr_1 dr_2 &= A\{L_1 w^{-1}[\mathfrak{S}^{\dagger 2}_{2,\lambda}(\psi_1, \psi_2)]\} \geqq A(\mathfrak{P}^2_{1,\lambda}) \\ &\geqq (1/2) \lg^2 \lambda + \lg 2 \cdot \lg (1/\lambda) - O[\lambda \lg (1/\lambda)]. \end{aligned} \tag{8.3.6}$$

Finally,

$$\begin{aligned} &\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\iint_{\mathfrak{S}^{\dagger 2}_2(\psi_1,\psi_2)} e(w_1, w_2) dr_1 dr_2\right] d\psi_1 d\psi_2 \\ &= \iiiint_{\mathfrak{B}^{\dagger}_2} [e(w_1, w_2)/|w_1 w_2|] d\omega_w \\ &= \iiiint_{w^{-1}(\mathfrak{B}^{\dagger}_2) \leqq \alpha_2} \left[\sum_{k=1}^{2} \frac{1}{|z_k|^2} \left| \frac{\partial w_1}{\partial z_{3-k}} \cdot \frac{\partial w_2}{\partial z_{3-k}} \right| \right] \frac{d\omega_z}{|w_1 w_2|} \\ &\leqq \alpha_2. \end{aligned} \tag{8.3.7}$$

by (5.1.2). Substituting from (8.3.6) into (8.3.2), integrating both sides of (8.3.2) with respect to ψ_k over the square $-\pi \leqq \psi_k < \pi$, and using (8.3.7), we obtain (8.3.1).

8.4. By using the method of Bergman [4, pp. 147–149], we now obtain an inequality which is a generalization of a result in one variable (see Spencer [3]). Suppose that the PT

$$\mathrm{t}:\quad \zeta_k = \zeta_k(\sigma_1, \sigma_2),$$

where $\sigma_k = \xi_k + i\eta_k$, transforms a domain $\mathfrak{G}_2$ into a domain $\mathfrak{G}_1$. Let $\mathfrak{Q}^2(\eta_1, \eta_2)$ be a set of points defined for each point (η_1, η_2) of a set $\mathfrak{A}^2$, and contained in the intersection $\mathfrak{G}_2 \cdot (\eta_k = \text{const.})$. Let

$$B(\eta_1, \eta_2) = \iint_{\mathfrak{Q}^2(\eta_1,\eta_2)} \left| \partial(\zeta_1, \zeta_2)/\partial(\sigma_1, \sigma_2) \right| d\xi_1 d\xi_2, \tag{8.4.1}$$

$$V_2 = \mathrm{V}\Big[\mathop{\mathfrak{S}}_{(\eta_1,\eta_2)\in\mathfrak{A}^2} \mathfrak{Q}^2(\eta_1, \eta_2)\Big], \qquad V_1 = \mathrm{V}\Big\{\mathrm{t}\Big[\mathop{\mathfrak{S}}_{(\eta_1,\eta_2)\in\mathfrak{A}^2} \mathfrak{Q}^2(\eta_1, \eta_2)\Big]\Big\}. \tag{8.4.2}$$

We then have

LEMMA 5.

$$\left[\iint_{\mathfrak{A}^2} B(\eta_1, \eta_2) d\eta_1 d\eta_2\right]^2 \leqq V_1 V_2. \tag{8.4.3}$$

The proof is trivial. In fact, writing

$$\mathrm{b} = \left| \partial(\zeta_1, \zeta_2)/\partial(\sigma_1, \sigma_2) \right|, \qquad d\omega = d\xi_1 d\xi_2 d\eta_1 d\eta_2, \qquad \mathfrak{J} = \mathop{\mathfrak{S}}_{(\eta_1,\eta_2)\in\mathfrak{A}^2} \mathfrak{Q}^2(\eta_1, \eta_2),$$

we have

$$\left[\iint_{\mathfrak{A}^2} B(\eta_1, \eta_2)\right]^2 = \left[\iiiint_{\mathfrak{J}} \mathrm{b}\, d\omega\right]^2$$

$$\leqq \left[\iiiint_{\mathfrak{J}} \mathrm{b}^2 d\omega\right] \cdot \left[\iiiint_{\mathfrak{J}} d\omega\right] = V_1 V_2,$$

by the inequality of Schwarz.

8.5. Let $\mathfrak{B}^{\dagger}_{2,\lambda_0}$ denote the portion of $\mathfrak{B}^{\dagger}_2$ for which $2|w_1 w_2| > \lambda_0$. We transform the domain $\mathbf{L}_2(\mathfrak{B}^{\dagger}_{2,\lambda_0})$ by means of the PT $\mathbf{L}_1 \mathbf{w}^{-1} \mathbf{L}_2^{-1}$, and apply Lemma 5 with $\mathfrak{G}_2 = \mathbf{L}_2(\mathfrak{B}^{\dagger}_{2,\lambda_0})$, $\mathfrak{G}_1 = \mathbf{L}_1 \mathbf{w}^{-1}(\mathfrak{B}^{\dagger}_{2,\lambda_0})$, $\eta_k = \psi_k$, $\xi_k = \lg r_k$, and

$$\mathfrak{A}^2 = \mathrm{E}[-\pi \leqq \psi_k < \pi].$$

We take

$$\mathfrak{Q}^2(\eta_1, \eta_2) = \mathfrak{Q}^2(\psi_1, \psi_2) = \mathbf{L}_2\big[\overset{\dagger 2}{\mathfrak{S}}_{2,\lambda_0}(\psi_1, \psi_2)\big].$$

Then by Lemma 4

$$\tag{8.5.1} \begin{aligned} \iint_{\Pi^2} B(\eta_1, \eta_2)d\eta_1 d\eta_2 &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B_{1,\lambda_0}(\psi_1, \psi_2)d\psi_1 d\psi_2 \\ &\geqq 2\pi^2 \lg^2 \lambda_0 + 4\pi^2 \lg 2 \cdot \lg (1/\lambda_0) - \alpha_2 \\ &\quad - O[\lambda_0 \lg (1/\lambda_0)], \end{aligned}$$

if $0<\lambda_0<2$ (as we suppose). Hence by Lemma 5

$$\tag{8.5.2} \begin{aligned} 2\pi^2 \lg^2 \lambda_0 + 4\pi^2 \lg 2 \cdot \lg (1/\lambda_0) - \alpha_2 - O[\lambda_0 \lg (1/\lambda_0)] \\ \leqq (V_1 \cdot V_2)^{1/2} \leqq (1/2)(V_1 + V_2), \end{aligned}$$

where

$$V_2 = \mathrm{V}[\mathbf{L}_2(\overset{\dagger}{\mathfrak{B}}_{2,\lambda_0})], \qquad V_1 = \mathrm{V}[\mathbf{L}_1\mathbf{w}^{-1}(\overset{\dagger}{\mathfrak{B}}_{2,\lambda_0})].$$

Let $\mathbf{L}$ denote the PT, $\mathrm{T}_k = \lg r_k$. The inequality (8.5.2) provides us with a lower bound in terms of λ_0 for the sum of $\mathrm{V}[\mathbf{L}_2(\mathfrak{B}^\dagger_{2,\lambda_0})]$ and $\mathrm{V}[\mathbf{L}_1\mathbf{w}^{-1}(\mathfrak{B}^\dagger_{2,\lambda_0})]$. To obtain the statement (5.1.3) of the theorem—that is, to obtain a lower bound for $\mathrm{V}(\mathfrak{B}^\dagger_{2,\lambda_0}) \cdot \mathrm{V}[\mathbf{w}^{-1}(\mathfrak{B}^\dagger_{2,\lambda_0})]$—we must rid (8.5.2) of the transformations $\mathbf{L}$. We note in this connection that the magnification by $\mathbf{L}$ of an element of volume at $(\lambda, \mu, \psi_1, \psi_2)$ is proportional to λ^{-2}. Therefore, to derive a lower bound for $\mathrm{V}(\mathfrak{B})$ from a lower bound for fixed $\mathrm{V}[\mathbf{L}(\mathfrak{B})]$, we must know something about the distribution of $\mathfrak{B}$ with respect to $\lambda=0$. This information is supplied here by the "mean valency" hypothesis on $\mathfrak{B}^\dagger_2$.

8.6. Since the w_k are regular in $\mathfrak{D}_1$, we have

$$\tag{8.6.1} \mathrm{V}(\overset{\dagger}{\mathfrak{B}}_{2,\lambda_0}) \leqq \mathrm{V}(\overset{\dagger}{\mathfrak{B}}_2) < \infty.$$

For brevity we now write

$$\tag{8.6.2} G(\Lambda) = (1/2)\int_0^\Lambda |\lg (2/\lambda)|\, d(\pi^2\lambda^2), \quad E(\Lambda) = G(\Lambda) + \alpha_1\Lambda^3 F(\Lambda),$$

where F is the function introduced in §4.5. ($G(\Lambda)$ is the volume of the portion of $\mathfrak{B}_0$ for which $2|w_1w_2| < \Lambda$.) Now we may plainly suppose (without loss of generality) that $E(\Lambda)$ is continuous. Then by (8.6.1) there exists a number Λ_2, $0<\Lambda_2<\infty$, such that

$$\tag{8.6.3} \mathrm{V}(\overset{\dagger}{\mathfrak{B}}_2) = E[\Lambda_2].$$

We let

$$\mathrm{n}^\dagger(r_1, r_2, \psi_1, \psi_2) = \mathrm{n}_{\mathfrak{B}_2^\dagger}(r_1, r_2, \psi_1, \psi_2)$$

be the number of times $\mathfrak{B}^\dagger_2$ covers the point $(r_1, r_2, \psi_1, \psi_2)$, $w_k = r_k e^{i\psi_k}$, and let

$$\overset{\dagger}{\mathrm{n}}_1(\lambda, \mu, \psi_1, \psi_2) = \overset{\dagger}{\mathrm{n}}(r_1, r_1, \psi_1, \psi_2),$$

where

(8.6.4) $$2r_1r_1 = \lambda,\ \mu = r_1^2 - r_2^2,\ dr_1dr_2 = d\mu d\lambda/4(\lambda^2 + \mu^2)^{1/2}.$$

Then

$$V_2 = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\iint_{E[2r_1r_2>\lambda_0]} \mathfrak{n}^{\dagger}(r_1, r_2, \psi_1, \psi_2)r_1^{-1}r_2^{-1}dr_1dr_2d\psi_1d\psi_2$$

(8.6.5) $$= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{\lambda_0}^{\infty}\int_{-\infty}^{\infty} \mathfrak{n}_1^{\dagger}(\lambda, \mu, \psi_1, \psi_2)\cdot\tfrac{1}{4}(\lambda^2 + \mu^2)^{-1/2}d\mu d(\lg \lambda^2)d\psi_1d\psi_2$$

$$= \int_{\lambda_0}^{\infty} \mathfrak{p}(\lambda)d(4\pi^2 \lg \lambda^2)$$

by (4.3.1) and (4.3.2). Next, integrating by parts,

$$\int_{\lambda_0}^{\infty}\mathfrak{p}(\lambda)d(\lg \lambda^2) = \int_{\lambda_0}^{\infty}\mathfrak{p}(\lambda)\cdot\lambda^{-2}d(\lambda^2) = \left[\Lambda^{-2}\int_{\lambda_0}^{\Lambda}\mathfrak{p}(\lambda)d(\lambda^2)\right]_{\Lambda=\lambda_0}^{\Lambda=\infty}$$

(8.6.6) $$+2\int_{\lambda_0}^{\infty}\left[\int_{\lambda_0}^{\Lambda}\mathfrak{p}(\lambda)d(\lambda^2)\right]\Lambda^{-3}d\Lambda$$

$$=2\int_{\lambda_0}^{\infty}\left[\int_{\lambda_0}^{\Lambda}\mathfrak{p}(\lambda)d(\lambda^2)\right]\Lambda^{-3}d\Lambda,$$

and hence, substituting from (8.6.6) into (8.6.5), we have that

(8.6.7) $$V_2 = 8\left\{\int_{\lambda_0}^{\Lambda_2} + \int_{\Lambda_2}^{\infty}\right\}\left[\int_{\lambda_0}^{\Lambda}\mathfrak{p}(\lambda)d(\pi^2\lambda^2)\right]\Lambda^{-3}d\Lambda = I_1 + I_2,$$

say. Now

(8.6.8) $$\int_0^{\Lambda}\mathfrak{p}(\lambda)d(\pi^2\lambda^2) \leqq E(\Lambda)$$

by (4.5.3), since $\mathfrak{B}_2^{\dagger}\in\mathfrak{T}(\alpha_1)$; and from the normalization (5.1.1) we see that

(8.6.9) $$\int_0^{\lambda_0}\mathfrak{p}(\lambda)d(\pi^2\lambda^2) = G(\lambda_0) + O(\lambda_0^3).$$

We suppose that $\lambda_0<\Lambda_2$, and then, using (8.6.8) and (8.6.9), we have

$$I_1 = 8\int_{\lambda_0}^{\Lambda_2}\left[\int_0^{\Lambda}\mathfrak{p}(\lambda)d(\pi^2\lambda^2) - \int_0^{\lambda_0}\mathfrak{p}(\lambda)d(\pi^2\lambda^2)\right]\Lambda^{-3}d\Lambda$$

(8.6.10) $$\leqq 8\int_{\lambda_0}^{\Lambda_2}[E(\Lambda) - G(\lambda_0)]\Lambda^{-3}d\Lambda + O(\lambda_0).$$

Also

(8.6.11)
$$I_2 = 8\int_{\Lambda_2}^{\infty}\left[\int_0^{\Lambda} p(\lambda)d(\pi^2\lambda^2) - \int_0^{\lambda_0} p(\lambda)d(\pi^2\lambda^2)\right]\Lambda^{-3}d\Lambda$$
$$\leqq 8\int_{\Lambda_2}^{\infty}[E(\Lambda_2) - G(\lambda_0)]\Lambda^{-3}d\Lambda + O(\lambda_0).$$

Let

(8.6.12)
$$G_1(\Lambda) = \begin{cases} G(\Lambda), & 0 \leqq \Lambda < \Lambda_2, \\ G(\Lambda_2), & \Lambda \geqq \Lambda_2, \end{cases} \qquad E_1(\Lambda) = \begin{cases} E(\Lambda), & 0 \leqq \Lambda < \Lambda_2, \\ E[\Lambda_2], & \Lambda \geqq \Lambda_2. \end{cases}$$

Then substituting from (8.6.10), (8.6.11) and (8.6.12) into (8.6.7) we obtain the inequality

(8.6.13)
$$V_2 \leqq 8\int_{\lambda_0}^{\infty}[E_1(\Lambda) - G_1(\lambda_0)]\Lambda^{-3}d\Lambda + O(\lambda_0)$$
$$\leqq 8\int_{\lambda_0}^{\infty}[G_1(\Lambda) - G_1(\lambda_0)]\Lambda^{-3}d\Lambda + 8\alpha_1 + O(\lambda_0)$$

by (4.5.1) and (4.5.2). A simple calculation now gives

(8.6.14)
$$8\int_{\lambda_0}^{\infty}[G_1(\Lambda) - G_1(\lambda_0)]\Lambda^{-3}d\Lambda = \frac{1}{2}\int_{\lambda_0}^{\Lambda_2}|\lg(2/\lambda)|\,d(4\pi^2\lg\lambda^2)$$
$$= 2\pi^2\lg^2\lambda_0 + 4\pi^2\lg 2\cdot\lg(1/\lambda_0)$$
$$+ \operatorname{sgn}(\Lambda_2 - 2)\cdot[2\pi^2\lg^2\Lambda_2 + 4\pi^2\lg 2.\lg(1/\Lambda_2)]$$
$$+ 2\pi^2\lg^2 2\cdot[1 + \operatorname{sgn}(\Lambda_2 - 2)],$$

where

$$\operatorname{sgn} a = \begin{cases} 1, & \text{if } a \geqq 0, \\ -1, & \text{if } a < 0. \end{cases}$$

and substituting from (8.6.14) into (8.6.13), we obtain

(8.6.15)
$$V_2 \leqq 2\pi^2\lg^2\lambda_0 + 4\pi^2\lg 2\cdot\lg(1/\lambda_0)$$
$$+ \operatorname{sgn}(\Lambda_2 - 2)\cdot[2\pi^2\lg^2\Lambda_2 + 4\pi^2\lg 2\cdot\lg(1/\Lambda_2)]$$
$$+ 2[1 + \operatorname{sgn}(\Lambda_2 - 2)]\pi^2\lg^2 2 + 8\alpha_1 + O(\lambda_0).$$

8.7. Since $\mathfrak{w}^{-1}(\mathfrak{B}_2^\dagger)$ is schlicht, and $\mathfrak{w}^{-1}(\mathfrak{B}_2^\dagger) \subset \mathfrak{B}_1 \subset \mathfrak{B}_0$, we see that $\mathfrak{w}^{-1}(\mathfrak{B}_2^\dagger)$ is mean one-valent with excess 0. Since

$$V[\mathfrak{w}^{-1}(\mathfrak{B}_2^\dagger)] \leqq V(\mathfrak{B}_1) = \pi^2,$$

we can choose a number Λ_1, $0 < \Lambda_1 \leqq 2$ such that

(8.7.1)
$$V[\mathfrak{w}^{-1}(\mathfrak{B}_2^\dagger)] = G(\Lambda_1),$$

and a calculation similar to that of §8.6 (but simpler) then gives

$$(8.7.2)\qquad \begin{aligned} V_1 \leqq\ & 2\pi^2 \lg^2 \lambda_0 + 4\pi^2 \lg 2\cdot \lg (1/\lambda_0) \\ & - 2\pi^2 \lg^2 \Lambda_1 - 4\pi^2 \lg 2\cdot \lg (1/\Lambda_1) + O[\lambda_0 \lg (1/\lambda_0)]. \end{aligned}$$

8.8. The theorem is now immediate. We suppose first that $0<\Lambda_2<2$. Then substituting from (8.6.15) and (8.7.2) into (8.5.2), and letting $\lambda_0\to 0$, we obtain

$$2\pi^2(\lg^2 \Lambda_1 + \lg^2 \Lambda_2) + 4\pi^2 \lg 2\cdot \lg (1/\Lambda_1\Lambda_2) \leqq 8\alpha_1 + 2\alpha_2,$$

and so *a fortiori*

$$\Lambda_1\Lambda_2 \geqq \exp [- (4\alpha_1 + \alpha_2)/2\pi^2 \lg 2].$$

If $\Lambda_2 \geqq 2$, we have similarly

$$2\pi^2(\lg^2 \Lambda_1 - \lg^2 \Lambda_2) + 4\pi^2 \lg 2\cdot \lg (\Lambda_2/\Lambda_1) \leqq 4\pi^2 \lg^2 2 + 8\alpha_1 + 2\alpha_2.$$

Either $\Lambda_1\Lambda_2 \geqq 1$, or $\Lambda_1 < 1/\Lambda_2$. In the latter case,

$$\lg^2 \Lambda_1 - \lg^2 \Lambda_2 \geqq 0,$$

and so (again *a fortiori*)

$$\Lambda_1 \geqq \tfrac{1}{2}\Lambda_2\cdot \exp [- (8\alpha_1 + 2\alpha_2)/4\pi^2 \lg 2] \geqq \exp [- (4\alpha_1 + \alpha_2)/2\pi^2 \lg 2].$$

Hence for all $\Lambda_2 > 0$,

$$(8.8.1)\qquad \begin{aligned} \Lambda_1\Lambda_2 &\geqq \min \{\exp [- (4\alpha_1 + \alpha_2)/2\pi^2 \lg 2], 1\} \\ &= \exp [- (4\alpha_1 + \alpha_2)/2\pi^2 \lg 2]. \end{aligned}$$

Finally

$$(8.8.2)\qquad V[\overset{-1}{w}(\overset{\dagger}{\mathfrak{B}}_2)]\cdot V[\overset{\dagger}{\mathfrak{B}}_2] = G(\Lambda_1)E(\Lambda_2) \geqq G(\Lambda_1)G(\Lambda_2).$$

If $0<\Lambda_2<2$, then

$$(8.8.3)\qquad \begin{aligned} G(\Lambda_1)G(\Lambda_2) &= \tfrac{1}{4}\pi^4\{[\lg 2 - \lg \Lambda_1]\Lambda_1^2 + \tfrac{1}{2}\Lambda_1^2\}\{[\lg 2 - \lg \Lambda_2]\Lambda_2^2 + \tfrac{1}{2}\Lambda_2^2\} \\ &\geqq (\pi^4/16)\Lambda_1^2\Lambda_2^2. \end{aligned}$$

Secondly, if $2 \leqq \Lambda_2 < 4$, then $G(\Lambda_2) \geqq \tfrac{1}{4}\pi^4\Lambda_2^2 \geqq \pi^2$, and so

$$(8.8.4)\qquad G(\Lambda_1)G(\Lambda_2) \geqq \pi^2 G(\Lambda_1) = (\pi^4/2)\{[\lg 2 - \lg \Lambda_1]\Lambda_1^2 + \tfrac{1}{2}\Lambda_1^2\} \geqq (\pi^4/64)\Lambda_1^2\Lambda_2^2.$$

Lastly, if $\Lambda_2 \geqq 4$,

$$(8.8.5)\qquad \begin{aligned} G(\Lambda_1)G(\Lambda_2) &= (\pi^4/4)\{[\lg 2 - \lg \Lambda_1]\Lambda_1^2 + \tfrac{1}{2}\Lambda_1^2\} \\ &\quad \cdot \{[\lg \Lambda_2 - \lg 2]\Lambda_2^2 + 4 - \tfrac{1}{2}\Lambda_2^2\} \\ &\geqq (\pi^4/8)\{[\lg 2 - \tfrac{1}{2}]\Lambda_2^2\}\cdot\{\Lambda_1^2\} \geqq (\pi^4/64)\Lambda_1^2\Lambda_2^2. \end{aligned}$$

The theorem (apart from the restriction that $n_{\mathfrak{B}_2}(0, \mu, \psi_1, \psi_2) \leqq 1$) follows from (8.8.1), (8.8.2), (8.8.3), (8.8.4) and (8.8.5).

9. The restriction that $n_{\mathfrak{B}_2}(0, \lambda, \psi_1, \psi_2) \leqq 1$

9.1. In §7 we supposed that

$$n_{\mathfrak{B}_2}(0, \lambda, \psi_1, \psi_2) \leqq 1. \tag{9.9.1}$$

We now indicate how this hypothesis may be removed. Since, however, the inclusion of all details would lengthen the paper considerably, we give only a sketch.

9.2. We proceed as follows. We define $\mathfrak{B}_2^\dagger$ as the subdomain of points $P(\lambda, \mu, \psi_1, \psi_2)$ of $\mathfrak{B}_2$ for which: (i) $\lambda > 0$; (ii) there exists a pair of numbers (ψ_1, ψ_2) and a path lying in

$$E[\arg w_k = \psi_k, (w_1, w_2) \in \mathfrak{B}_2, |w_k| > 0]$$

which connects (in the sense of §7) P to a neighborhood of the point $\mathbf{w}(0, 0)$. Let $\mathfrak{B}_2^{\dagger\dagger}$ be the subdomain of $\mathfrak{B}_2$ any point P of which can be connected with $\mathbf{w}(0, 0)$ along a path the transform of which in $\mathfrak{B}_1$ is contained in

$$E[\arg w_k(z_1, z_2) = \psi_k, |z_k| < 1] - E[|z_1| = 0] - E[|z_2| = 0].$$

Then $\mathfrak{B}_2^\dagger \subset \mathfrak{B}_2^{\dagger\dagger}$. We have

Lemma 6. *If $\mathfrak{B}_2^\dagger \in \mathfrak{T}(\alpha_1)$, then $\mathfrak{B}_2^\dagger$ and $\mathfrak{B}_2^{\dagger\dagger}$ are identical.*

In fact, let $n^{\dagger\dagger}(\lambda, \mu, \psi_1, \psi_2)$ be the number of times $\mathfrak{B}_2^{\dagger\dagger}$ covers the point $(\lambda, \mu, \psi_1, \psi_2)$. Then Lemma 6 is plainly equivalent to the assertion that

$$n^{\dagger\dagger}(0, \mu, \psi_1, \psi_2) \leqq 1. \tag{9.2.1}$$

Suppose now that a point P_0 on $\lambda = 0$ were covered by $\mathfrak{B}_2^{\dagger\dagger}$ twice. Then there is a path

$$\mathbf{r}^1 = E[r_1(s)e^{i\psi_1^0}, r_2(s)e^{i\psi_2^0}, 0 \leqq s \leqq 1],$$

with

$$[r_1(0)e^{i\psi_1}, r_2(0)e^{i\psi_2}] = P_0 \quad [r_1(1), r_2(1)] = \mathbf{w}(0, 0),$$

lying, except for its end points, in the portion of $\mathfrak{B}_2$ for which $\lambda > 0$. Hence there is a $\mathrm{P} > 0$ for which the "tube"

$$\mathfrak{R} = \underset{(\omega_1, \omega_2) \in \mathbf{r}^1}{\mathfrak{S}} [|w_1 - \omega_1|^2 + |w_2 - \omega_2|^2 \leqq \mathrm{P}^2]$$

is contained in $\mathfrak{B}_2$. Since $\mathfrak{R}$ lies in a finite region of space, there is an ϵ, $\epsilon > 0$ such that the set

$$E[(r_1(s) + \epsilon_1)e^{i(\psi_1 + \epsilon_2)}, (r_2(s) + \epsilon_3)e^{i(\psi_2 + \epsilon_4)}, |\epsilon_j| < \epsilon]$$

is contained in $\mathfrak{R}$. If

$$n^\dagger(\lambda, \mu, \psi_1, \psi_2)$$

is the number of times $\mathfrak{B}_2^\dagger$ covers the point $(\lambda, \mu, \psi_1, \psi_2)$, we have, therefore

$$\mathfrak{n}^\dagger(\lambda, \mu, \psi_1, \psi_2) > 1 \tag{9.2.2}$$

in a four-dimensional domain

$$\begin{aligned}\mathfrak{N}(\lambda_1) = E[0 < \lambda < \overset{(1)}{\lambda}, \overset{(1)}{\mu} < \mu < \overset{(2)}{\mu}, \\ \overset{(1)}{\psi_1} < \psi_1 < \overset{(2)}{\psi_1}, \overset{(1)}{\psi_2} < \psi_2 < \overset{(2)}{\psi_2}].\end{aligned} \tag{9.2.3}$$

From the normalization (5.1.1) we now have

$$\begin{aligned} V[\mathfrak{B}^\dagger(\Lambda)] &= \frac{1}{16}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{0}^{\Lambda}\int_{-\infty}^{\infty}\mathfrak{n}^\dagger(\lambda, \mu, \psi_1, \psi_2)\frac{d\mu}{(\lambda^2+\mu^2)^{1/2}}d(\lambda^2)d\psi_1 d\psi_2 \\ &\geqq \frac{1}{16}\int_{-\pi}^{\pi}d\psi_1\int_{-\pi}^{\pi}d\psi_2\int_{0}^{\Lambda}d(\lambda^2)\int_{-(1-\lambda^2/4)}^{(1-\lambda^2/4)}(\lambda^2+\mu^2)^{-1/2}d\mu \\ &\quad + \frac{1}{16}\iiiint_{\mathfrak{N}(\lambda)}(\lambda^2+\mu^2)^{-1/2}d\mu d(\lambda^2)d\psi_1 d\psi_2 + O(\Lambda^3) \\ &= V[\mathfrak{B}_0(\Lambda)] + \frac{1}{16}\iiiint_{\mathfrak{N}(\lambda)}(\lambda^2+\mu^2)^{-1/2}d\mu d(\lambda^2)d\psi_1 d\psi_2 \\ &\quad + O(\Lambda^3) > V[\mathfrak{B}_0(\Lambda)] + \alpha_1\Lambda^3 F(\Lambda), \end{aligned} \tag{9.2.4}$$

for a sequence of Λ tending to zero. This contradicts (4.5.3), and proves Lemma 6.

9.3. Finally, by Lemma 6 we see that for the mapping (6.2.1) there exists a $\delta = \delta(\rho)$ such that

$$\mathfrak{n}^\dagger(\lambda, \mu, \psi_1, \psi_2) \leqq 1, \tag{9.3.1}$$

if $0 < \lambda < \delta$.

We now apply the method of §7 to $\mathfrak{S}_2^{\dagger 2}(\psi_1, \psi_2)$ *without introducing* $\mathfrak{S}_2^2(\psi_1, \psi_2)$. Apart from minor complications, the proof then proceeds substantially as written out above.

Bibliography

Stefan Bergman

1. *Über eine in der Theorie der Funktionen von zwei komplexen Veränderlichen auftretende unitäre Geometrie*, Proceedings, Akademie van Wetenschappen, Amsterdam, vol. 36 (1933), pp. 307–313.
2. *Sur quelques propriétés des transformations par un couple des fonctions de deux variables complexes*, Rendiconti delle R. Accademia Nazionale dei Lincei, (6), vol. 19 (1934), pp. 474–478.
3. *Über eine Abschätzung von meromorphen Funktionen zweier komplexen Veränderlichen in Bereichen mit ausgezeichneter Randfläche*, Travaux de l'Institut Mathématique de Tbilissi, vol. 1 (1937), pp. 187–204.
4. *Theory of pseudo-conformal transformations and its connection with differential geometry*, Notes of lectures delivered at the Massachusetts Institute of Technology, 1939–1940 (available at the Brown University library).
5. *Lectures on conformal mapping of multiply-connected domains* (Partial Differential Equations, Advanced Topics), Publication of Brown University, 1941.

6. *Sur les Fonctions Orthogonales de Plusieurs Variables Complexes avec les Applications à la Théorie des Fonctions Analytiques*, Interscience Publishers, New York, 1941.

A. Bermant

1. *Remarque sur le lemme de Schwarz*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 207 (1938), pp. 31–33.

G. M. Golusin

1. *Einige Überdeckungssatze für die im Kreise regulären Funktionen*, Recueil Mathématique, (n.s.), vol. 2(44) (1937), pp. 617–619.

W. F. Osgood

1. *Lehrbuch der Funktionentheorie*, vol. 2, part 1.

D. C. Spencer

1. *On finitely mean valent functions*, Proceedings of the London Mathematical Society, (2), vol. 47 (1941), pp. 201–211.
2. *On an inequality of Grunsky*, Proceedings of the National Academy of Sciences, vol. 26 (1940), pp. 616–621.
3. *On a theorem of Rengel*, Journal of Mathematics and Physics, vol. 20 (1941), pp. 118–123.

Brown University,
Providence, R.I.
Massachusetts Institute of Technology,
Cambridge, Mass.

LEBESGUE THEORY ON A BOOLEAN ALGEBRA

BY

JOHN M. H. OLMSTED

Introduction and definitions

Associated with a real-valued point function everywhere finite is the set-valued function

$$E(\alpha) = \underset{x}{E}[f(x) > \alpha].$$

The function $E(\alpha)$ in turn determines $f(x)$:

$$f(x) = \sup \underset{\alpha}{E}[x \in E(\alpha)].$$

A necessary and sufficient set of conditions that a function, $E(\alpha)$, from real numbers to subsets of a set S, be associated, as above, with a real-valued point function everywhere finite is

1. $E(\alpha) \downarrow$ *as* $\alpha \uparrow$;
2. $\sum_{n=1}^{\infty} E(-n) = S, \prod_{n=1}^{\infty} E(n) = 0$;
3. $\sum_{n=1}^{\infty} E(\alpha + 1/n) = E(\alpha)$ *for every* α.

It is possible, therefore, to consider $f(x)$ and $E(\alpha)$ as different aspects of the same function, one more natural for algebraic combinations of functions, and the other receiving emphasis in a theory of integration. However, one may restrict oneself entirely to the second aspect. For example, if

$$A(\alpha) = \underset{x}{E}[f(x) > \alpha],$$

$$B(\alpha) = \underset{x}{E}[g(x) > \alpha],$$

$$C(\alpha) = \underset{x}{E}[f(x) + g(x) > \alpha],$$

then

$$C(\alpha) = \sum_{\beta} \prod [A(\beta), B(\alpha - \beta)]$$

where the summation is formed with respect to any dense set of real numbers β.

This fact frees us altogether from the point aspect of functions and enables us to extend our attention to functions whose values are elements of a Boolean algebra and which cannot in general be regarded as point functions. If there is a measure function we may have "almost everywhere" instead of

Presented to the Society, September 11, 1940; received by the editors December 9, 1940.

"everywhere" finite, and consider the new Boolean algebra of equivalence classes of elements differing by elements of measure zero.

Chapters one and two deal with algebraic manipulations of functions and extensions of partially ordered linear spaces with element 1 as defined by Freudenthal [1](1). It is proved that any such space can be extended to a partially ordered space of the same type which is also a ring, the multiplicative unit of which is the element 1. Spaces analogous to the space M of bounded measurable functions and, in the case of a measure on the Boolean algebra, to the Lebesgue L^p spaces are treated in Chapters three and four. In Chapter five the notions of absolute continuity, indefinite integral, and derivative are studied. It is shown that in terms of these concepts the Nikodym theorem (cf. Saks [2, chap. 1, §14]) concerning absolutely continuous functions is again available. Chapter six contains remarks on conjugate spaces, regularity for a partially ordered linear space as defined by Kantorovitch [3], and spaces of the type defined by Orlicz [4].

Spaces of functions based on a Boolean algebra have been considered from a different approach by Carathéodory [5]. The present paper rests particularly on the work of Kantorovitch [3] and Freudenthal [1].

Terminology will be in large part that used by Birkhoff in his *Lattice Theory* [8]. Definitions given there will not be repeated here. The symbols $\bigvee$ and $\bigwedge$ will be used to represent the "sup" and "inf," respectively, of a collection of elements of a lattice. The notation of Kantorovitch, $a^- = \bigvee(-a, 0)$, where a is an element of a vector lattice, will be retained. A vector lattice will also be called a K-space (Kantorovitch space). If a vector lattice has an element 1, that is, an element $1>0$ with the property that

$$\bigwedge(a, 1) = 0 \rightarrow a = 0,$$

it will also be called an F-space (Freudenthal space). Real numbers will in general be denoted by Greek letters, integers generally by m, n. The letters, p, q will also on occasion be used for real numbers, generally $\geqq 1$. Latin letters will generally be reserved for elements of a lattice. We shall frequently write $f(\xi)$ to mean the function $f(\xi)$ itself rather than its value at a point ξ. It will be also convenient to denote the function by the single letter f if there is no danger of confusion.

Throughout this paper the following fact holds. If the Boolean algebra under consideration is an algebra of point sets (cf. Stone [6], Wecken [7]), the theory reduces to the classical case.

I. The space Ω

Notation. *Let B be a σ-complete Boolean algebra. Denote by Ω the set of all functions, $f(\alpha)$, from real numbers to elements of B, satisfying the conditions:*

(1) Numbers in brackets indicate references at the end of the paper.

1. $f(\alpha)\downarrow$ as $\alpha\uparrow$,
2a. $\bigvee_\alpha f(\alpha)=1$,
2b. $\bigwedge_\alpha f(\alpha)=0$,
3. $\bigvee_{\beta>\alpha} f(\beta)=f(\alpha)$ *for every* α.

Remark. Because of the monotoneity condition 1, all $\bigvee$ and $\bigwedge$ indicated exist.

With proper definitions of ordering, addition, and multiplication by real numbers, the set Ω can be regarded as a σ-complete F-space. The symbol Ω will be used to designate this space as well as the set of its elements. We shall now define and discuss the operations in the space Ω.

1. **Ordering.** We state the following definition.

DEFINITION. *If $f(\xi)$ and $g(\xi)$ are elements of Ω, we define $f(\xi)\leqq g(\xi)$ to mean that $f(\xi)\leqq g(\xi)$ for every ξ, $f(\xi)=g(\xi)$ to mean that $f(\xi)=g(\xi)$ for every ξ, and $f(\xi)<g(\xi)$ to mean that $f(\xi)\leqq g(\xi)$, but $f(\xi)\neq g(\xi)$.*

Remark. Under this definition, Ω becomes a partially ordered set.

THEOREM 1.1. *If f_n, $g\in\Omega$, $f_n\leqq g$, then $f=\bigvee_n f_n$ exists and*

$$f(\xi) = \bigvee_n f_n(\xi).$$

If f_n, $g\in\Omega$, $f_n\geqq g$, then $f=\bigwedge_n f_n$ exists and

$$f(\xi) = \bigvee_{m=1}^{\infty} \bigwedge_n f_n(\xi + 1/m).$$

If $f_1, \cdots, f_n\in\Omega$, then $h=\bigvee(f_1, \cdots, f_n)$ and $f=\bigwedge(f_1, \cdots, f_n)$ exist and

$$f(\xi) = \bigwedge (f_1(\xi), \cdots, f_n(\xi)).$$

Proof is omitted.

2. **Addition.** We state the following definition.

DEFINITION. *If $f(\xi)$ and $g(\xi)$ are elements of Ω, we define their sum, $h(\xi)=f(\xi)+g(\xi)$ to be the function*

$$h(\xi) = \bigvee_\beta \bigwedge [f(\beta), g(\xi - \beta)],$$

where the "sup" *is formed with respect to any countable dense set of real numbers β.*

To speak of addition as thus defined we must show that the values of $h(\xi)$ are independent of the dense set chosen and that the function so defined belongs to Ω. Let us first observe that due to the normalization condition 3 a function in Ω is completely determined if its values over a dense set of real numbers are known. We state two lemmas:

Lemma 1. *If $a_n \uparrow$ and $b_n \uparrow$ are two increasing sequences of elements of B, then*

$$\bigvee_n \bigwedge = (a_n, b_n) = \bigwedge \left(\bigvee_m a_m, \bigvee_n b_n\right),$$

Proof is omitted.

Lemma 2. *If $h = f + g$, as defined above, then for any ξ the value $h(\xi)$ is independent of the choice of the dense set β.*

To verify this it will be sufficient to show that for any ρ

$$\bigvee_\beta \bigwedge [f(\beta), g(\xi - \beta)] \geqq \bigwedge [f(\rho), g(\xi - \rho)].$$

The left-hand side is greater than or equal to

$$\bigvee_{\beta > \rho} \bigwedge [f(\beta), g(\xi - \beta)] \geqq \bigvee_{\beta > \rho} \bigwedge [f(\beta), g(\xi - \rho)],$$

which by the distributive law for B is equal to

$$\bigwedge \left[\bigvee_{\beta > \rho} f(\beta), g(\xi - \rho)\right] = \bigwedge [f(\rho), g(\xi - \rho)].$$

Theorem 1.2. *If f and $g \in \Omega$, then $h = f + g \in \Omega$.*

Proof. We shall verify the second of the three conditions.

Ad 2a: Form the sum with respect to a dense set $\{\rho\}$. Then

$$\begin{aligned}\bigvee_n \bigvee_\rho \bigwedge [f(\rho), g(-n-\rho)] &= \bigvee_\rho \bigvee_n \bigwedge [f(\rho), g(-n-\rho)] \\ &= \bigvee_\rho \left[\bigwedge \left[f(\rho), \bigvee_n g(-n-\rho)\right]\right] \\ &= \bigvee_\rho [\bigwedge [f(\rho), 1]] = 1.\end{aligned}$$

Ad 2b: We have

$$\begin{aligned}&\bigwedge_n \bigvee_\rho \bigwedge [f(\rho), g(n-\rho)] \\ &\quad = \bigwedge_n \bigvee \left[\bigvee_{\rho > n} \bigwedge [f(\rho), g(2n-\rho)], \bigvee_{\rho < n} \bigwedge [f(\rho), g(2n-\rho)]\right] \\ &\quad \leqq \bigwedge_n \bigvee \left[\bigvee_{\rho > n} \bigwedge [f(n), 1], \bigvee_{\rho < n} \bigwedge [1, g(n)]\right] = \bigwedge_n \bigvee [f(n), g(n)].\end{aligned}$$

Because of the monotoneity of $f(\xi)$ and $g(\xi)$ we can apply Lemma 1. The last expression above is therefore equal to zero.

Theorem 1.3. *Addition is commutative.*

Proof is omitted.

Theorem 1.4. *Addition is associative.*

Proof is based on the associative and distributive laws for B.

3. **Multiplication by real numbers.** We give first a definition.

DEFINITION. *If $\alpha>0$, $g=\alpha f$ is the function*

$$g(\xi) = f(\xi/\alpha).$$

The function $g=-f$ is given by

$$g(\xi) = 1 - \bigwedge_{\rho<-\xi} f(\rho) = 1 - \bigwedge_{n=1}^{\infty} f(-\xi - 1/n).$$

The function $g=0=0f$ is given by

$$g(\xi) = 0(\xi) = \begin{cases} 1, & \xi < 0, \\ 0, & 0 \leqq \xi. \end{cases}$$

If $\alpha<0$, $g=\alpha f$ is the function $-[(-\alpha)f]$.

There is no difficulty in showing in each case that the defined function belongs to Ω.

THEOREM 1.5. *1. $\alpha(\beta f)=(\alpha\beta)f$, 2. $(\alpha+\beta)f=\alpha f+\beta f$, 3. $\alpha(f+g)=\alpha f+\alpha g$, for all real numbers, α, β, and for any elements f, g of Ω.*

Proof. Ad 1: The following statements, from which the general form follows, are easily proved:

(1) $-(-f)=f$,
(2) $-(\alpha f)=\alpha(-f)$, if $\alpha>0$,
(3) $1\cdot f=f$,
(4) $(-1)f=-f$,
(5) $-(0)=0$, where $0=0(\xi)$,
(6) $\alpha(0(\xi))=0(\xi)$ for every α,
(7) $\alpha(\beta f)=(\alpha\beta)f$, if $\alpha=0$, or if $\beta=0$,
(8) $\alpha(\beta f)=(\alpha\beta)f$, if $\alpha>0$, and $\beta>0$.

Ad 2: Proof involves such preliminaries as showing that $f+0=f$ and that $f+(-f)=0$, for any f. The only one involving any difficulty is the proof that $f+(-f)=0$. For any real number ξ

$$[f+(-f)](\xi) = \bigvee_{\rho} \wedge \left[1 - \bigwedge_{\beta<-\rho} f(\beta), f(\xi-\rho)\right].$$

If $\xi\geqq 0$, there is no difficulty in showing that this expression is equal to zero. Now let us assume that $\xi<0$, and show that the above expression is equal to 1. We know that it is greater than or equal to

$$\bigvee_{\rho} \wedge \left[1 - f(\xi/2-\rho), f(\xi-\rho)\right].$$

Setting $2\alpha=-\xi$, $\alpha>0$, and letting ρ run over the infinite range of values $n\alpha$, for all integers n (we can add these values to the dense set of ρ without alter-

ing the supremum), we see that the expression above is greater than or equal to

$$\begin{aligned}\bigvee [\cdots \wedge [f(-n\alpha), 1 - f(-(n-1)\alpha)], \cdots,\\ \wedge [f(-\alpha), 1 - f(0)], \wedge [f(0), 1 - f(\alpha)], \cdots]\\ = \bigvee_n \bigvee [\wedge [f(-n\alpha), 1 - f(-(n-1)\alpha)], \cdots,\\ \wedge [f((n-1)\alpha), 1 - f(n\alpha)]].\end{aligned}$$

Iterating the relation $\vee[\wedge(a, 1-b), \wedge(b, 1-c)] = \wedge(a, 1-c)$, where $a \geqq b \geqq c$, we obtain the result that the last expression above is equal to

$$\bigvee_n \wedge [f(-n\alpha), 1 - f(n\alpha)] = \wedge [\bigvee_n f(-n\alpha), 1 - \bigwedge_n f(n\alpha)] = \wedge (1, 1) = 1.$$

Ad 3: Proof is not difficult.

This completes the proof of Theorem 1.5.

It follows immediately from the definition of ordering that a necessary and sufficient condition that an element $f(\xi)$ of Ω be $\geqq 0$ is that $f(\xi) = 1$ for $\xi < 0$. It is now easily proved that if $f > 0$, $g > 0$, $\alpha > 0$, then $f + g > 0$, $\alpha f > 0$. Furthermore, if $f \in \Omega$, then $f^+ \in \Omega$. It is given by

$$f^+(\xi) = \begin{cases} 1, & \xi < 0, \\ f(\xi), & \xi \geqq 0. \end{cases}$$

THEOREM 1.6. *The space Ω is a σ-complete F-space.*

Proof. We have only to show that there exists an element $1 \in \Omega$. Such an element is the function

$$1(\xi) = \begin{cases} 1, & \xi < 1, \\ 0, & \xi \geqq 1. \end{cases}$$

Corresponding to any element a of B we define the function $a(\xi)$:

$$a(\xi) = \begin{cases} 1, & \xi < 0, \\ a, & 0 \leqq \xi < 1, \\ 0, & 1 \leqq \xi. \end{cases}$$

If a is the element 0 or 1 of the Boolean algebra, the corresponding function is the element 0 or 1, previously defined, of the space Ω. In case there is no danger of confusion, we shall use a to denote the function itself. We shall borrow a term from real variable theory and call these functions "characteristic functions."

DEFINITION. *Two elements, a and b, of a Boolean algebra are "disjoint" if $\wedge(a, b) = 0$. Two elements a and b, of a vector lattice are "disjoint" if $\wedge(|a|, |b|) = 0$.*

RING PROPERTY OF Ω

The space Ω bears a closer resemblance to a space of point functions than the general σ-complete F-space, for it admits the introduction of a multiplication and a raising of positive elements to powers, satisfying the usual laws.

NOTATION. *We denote by Ω^+ the elements of Ω which are $\geqq 0$.*

DEFINITION. *If f and g are elements of Ω^+, we define their product, $h=fg$, to be the function*

$$h(\xi) = \begin{cases} 1, & \xi < 0, \\ \bigvee_{\rho>0} \bigwedge [f(\rho), g(\xi/\rho)], & \xi \geqq 0, \end{cases}$$

where the least upper bound is formed with respect to any countable set of points, ρ, dense in the set of positive real numbers.

As in the case of sums, the values of $h(\xi)$ are independent of the dense set chosen. Similarly we have

THEOREM 1.7. *If f, $g \in \Omega^+$, then $fg \in \Omega^+$.*

Proof is omitted.

Let us observe that if $f \geqq 0$, $g \geqq 0$, then a necessary and sufficient condition that $fg=0$ is that $\bigwedge(f, g)=0$, that is,

$$\bigwedge [f(0), g(0)]=0.$$

DEFINITION. *If $f \in \Omega^+$ and if p is a positive real number, then $g=f^p$ is defined:*

$$g(\xi) = \begin{cases} 1, & \xi < 0, \\ f(\xi^{1/p}), & \xi \geqq 0. \end{cases}$$

We now state the following theorem about functions $\in \Omega^+$:

THEOREM 1.8. 1. $fg=gf$, 2. $(fg)h=f(gh)$, 3. $f(g+h)=fg+fh$, 4. $f0=0$, 5. $f \cdot 1=f$, 6. *if* $\alpha>0$, $(\alpha f)(g)=\alpha(fg)$, 7. $f^pf^q=f^{p+q}$, *for* $p>0$, $q>0$, 8. $f^pg^p=(fg)^p$, *for* $p>0$, 9. *if* $fg=0$, *then* $(f+g)^p=f^p+g^p$, 10. $a^p=a$, *for* $p>0$, 11. $\bigvee[f^p, g^p] = [\bigvee(f, g)]^p$, $\bigwedge[f^p, g^p]=[\bigwedge(f, g)]^p$, *for* $p>0$, 12. $\bigvee_n(f_n^p)=(\bigvee_n f_n)^p$, *for* $p>0$.

Proof is omitted.

DEFINITION. *For arbitrary elements of Ω we define*

$$fg = f^+g^+ + f^-g^- - f^+g^- - f^-g^+.$$

LEMMA. $(fg)^+=f^+g^++f^-g^-$, $(fg)^-=f^+g^-+f^-g^+$.

Proof. The statement of the lemma is equivalent to stating

$$\bigwedge (f^+g^+ + f^-g^-, f^+g^- + f^-g^+) = 0,$$

which is equivalent to saying

$$(f^+g^+ + f^-g^-)(f^+g^- + f^-g^+) = 0,$$

and by the distributive, associative, and commutative laws this is true because $f^+f^- = g^+g^- = 0$.

We are now in a position to state a theorem regarding multiplication of arbitrary elements of Ω:

THEOREM 1.9. 1. $fg=gf$, 2. $(fg)h=f(gh)$, 3. $f(g+h)=fg+fh$, 4. $f\cdot 0=0$, 5. $f\cdot 1=f$, 6. $(\alpha f)g=\alpha(fg)$, 7. $|fg|=|f|\cdot|g|$, 8. $|\alpha f|=|\alpha|\cdot|f|$.

Proof is omitted.

DEFINITION. *If $f\geqq 0$, we define*

$$f_a(\xi) = \begin{cases} 1, & \xi < 0, \\ \wedge\,[f(\xi), a], & \xi \geqq 0. \end{cases}$$

For arbitrary $f\in\Omega$, we define

$$f_a = (f^+)_a - (f^-)_a.$$

One could now prove such statements as the following:

$$(f^+)_a = (f_a)^+;\ (f+g)_a = f_a + g_a;\ (f_a)_b = (f_b)_a = f_{\wedge(a,b)};\ (fg)_a = f_a g_a;$$

if $f\geqq 0$, $f_a = \vee_n \wedge(f, na)$. Results of this nature will be used after the notion of integral has been introduced.

DEFINITION. *A function $f(\xi)$ is "bounded above" if there exists an α such that $f(\alpha)=0$; $f(\xi)$ is "bounded below" if there exists a β such that $f(\beta)=1$; $f(\xi)$ is "bounded" if it is bounded above and bounded below.*

DEFINITION. *A simple function is one which takes on only a finite number of values. A super-simple function is a simple function which is greater than or equal to zero and which takes on at most one value other than* 0 *and* 1.

THEOREM 1.10. *Any positive simple function can be represented as a sum of disjoint super-simple functions. Any simple function can be represented as a linear combination of disjoint "characteristic functions."*

Proof. We prove the first part. Let the given positive simple function be

$$s(\xi) = \begin{cases} 1, & \xi < 0, \\ a_1, & 0 \leqq \xi < \alpha_1, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ , \\ a_n, & \alpha_{n-1} \leqq \xi < \alpha_n, \\ 0, & \alpha_n \leqq \xi. \end{cases}$$

Define

$$u(\xi) = \begin{cases} 1, & \xi < 0, \\ a_2, & 0 \leqq \xi < \alpha_2, \\ \cdots\cdots\cdots, & \\ a_n, & \alpha_{n-1} \leqq \xi < \alpha_n, \\ 0, & \alpha_n \leqq \xi; \end{cases} \qquad t(\xi) = \begin{cases} 1, & \xi < 0, \\ a_1 - a_2, & 0 \leqq \xi < \alpha_1, \\ 0, & \alpha_1 \leqq \xi; \end{cases}$$

and $v(\xi) = u(\xi) + t(\xi)$. Then

$$v(\xi) = \bigvee \begin{cases} t(\xi), \ \wedge [a_2, t(\xi - \alpha_2)], \cdots, \wedge [a_n, t(\xi - \alpha_n)], \\ u(\xi), \ \wedge [u(\xi - \alpha_1), a_1 - a_2], \end{cases}$$

and by direct comparison $v(\xi) = s(\xi)$ for all ξ. Therefore a positive function with at most $n+2$ values can be represented as the sum of a positive function with at most $n+1$ values and a super-simple function, for any positive integer n. Induction completes the proof.

Given any two simple functions, s and t, we may write them as linear combinations of the same characteristic functions, that is, in the form

$$s = \alpha_1 a_1 + \cdots + \alpha_n a_n, \qquad t = \beta_1 a_1 + \cdots + \beta_n a_n.$$

Using the distributive laws, we can now write:

$$s + t = (\alpha_1 + \beta_1)a_1 + \cdots + (\alpha_n + \beta_n)a_n,$$
$$st = (\alpha_1\beta_1)a_1 + \cdots + (\alpha_n\beta_n)a_n.$$

THEOREM 1.11. *The simple functions of Ω form an F-space possessing the ring property.*

Proof is omitted.

Since Ω is a σ-complete vector lattice it admits the introduction of a limit (cf. Birkhoff [8, chap. 2]), with respect to which we can state:

THEOREM 1.12. *The space of simple functions is dense, in the sense of this limit, in the space Ω.*

Proof. Suppose $f \geqq 0$. Take any countable dense set of positive real numbers, ρ_n, and letting $0 < \beta_1 < \cdots < \beta_n$ be the rearrangement of $\rho_1, \cdots, \rho_n$ according to magnitude, define

$$f_n(\xi) = \begin{cases} 1, & \xi < 0, \\ f(\beta_1), & 0 \leqq \xi < \beta_1, \\ \cdots\cdots\cdots\cdots, & \\ f(\beta_n), & \beta_{n-1} \leqq \xi < \beta_n, \\ 0, & \beta_n \leqq \xi. \end{cases}$$

Then f_n is an increasing sequence of simple functions and

$$\bigvee_n f_n(\xi) = \bigvee_{\rho>\xi} f(\rho) = f(\xi).$$

Finally, if f is an arbitrary function $\in\Omega$, write $f=f^+-f^-$, and let $\bigvee_n g_n=f^+$, $\bigvee_n h_n=f^-$, $f_n=g_n-h_n$. Then f_n is a simple function, and

$$\lim f_n = \lim g_n - \lim h_n = f^+ - f^- = f.$$

THEOREM 1.13. *Let $f, g\in\Omega^+$, and let s_n and t_n be simple functions $\in\Omega^+$. If $s_n \uparrow f$, $t_n \uparrow g$, then $s_n+t_n \uparrow f+g$ and $s_nt_n \uparrow fg$.*

Proof. The first statement is proved by Birkhoff [8, p. 112]. The second statement of the theorem follows from Lemma 1 preceding Theorem 1.2.

II. EXTENSIONS OF SPACES WITH ELEMENT 1

Let F be a σ-complete vector lattice with element 1. As proved by Freudenthal [1], there is a subset B of elements of F, between 0 and 1, which form a σ-complete Boolean algebra, where the partial ordering of B is the same as that of F. The subset B is defined as follows: an element e of F belongs to B if

$$\bigwedge (e, 1 - e) = 0.$$

To an arbitrary element a of F can be made to correspond an element of B, namely the greatest lower bound of all elements of B which are greater than or equal to $\bigwedge(1, a)$. This "inf" exists, since it can be computed as the countable "sup:"

$$e(a) = \bigvee_{n=1}^{\infty} \bigwedge [(na)^+, 1].$$

The following relation holds for any element a of F: $e(a)+e(-a)\leqq 1$.

Now let c be a *fixed* element of F. Then there is defined the function from real numbers to elements of B:

$$f(\alpha) = e_\alpha = e(c - \alpha).$$

Freudenthal proves that c not only determines the function $f(\alpha)$, but that if $f(\alpha)$ is given by an element c, it in turn determines c, by a Lebesgue-Stieltjes integral representation.

Let Ω be the space of functions, as defined in Chapter I, determined by B. Considering Ω as a set of functions, we have

THEOREM 2.1. *If $c\in F$, and $f(\alpha)=e(c-\alpha)$, then $f(\alpha)\in\Omega$.*

Proof. We prove part two, observing first that the bounds indicated exist because of the monotoneity of $f(\alpha)$.

Ad 2b: We establish first the inequality $\wedge(mc-mn,\ n)\leqq c$, where $c\geqq 0$, and m and n are integers $\geqq 0$. Represent the left side of the inequality by x. Then $x\leqq n$, and therefore $-mn\leqq -mx$. Also $x\leqq mc-mn$ and therefore $x\leqq mc-mx$. That is, $(m+1)x\leqq mc\leqq(m+1)c$, or $x\leqq c$, as was to be proved. Now, since $0\leqq c$,

$$\vee\ [\wedge\ (mc - mn,\ n),\ 0] \leqq c,$$

or, by the distributive law for F,

$$\wedge\ [\vee\ (mc - mn,\ 0),\ n] \leqq c.$$

Therefore for every m and $n\geqq 0$

$$n \wedge [m(c - n)^+,\ 1] = \wedge\ [mn(c - n)^+,\ n] \leqq c,$$

and so for every $n\geqq 0$

$$n \bigvee_{m=1}^{\infty} \wedge\ [m(c - n)^+,\ 1] = ne(c - n) \leqq c.$$

Therefore for any c and for every integer $n\geqq 0$,

$$ne(c - n) \leqq ne(c^+ - n) \leqq c^+.$$

Letting $e=\wedge_n e(c-n)$, we have $ne\leqq c^+$ for every $n\geqq 0$. But this implies that $e=0$.

Ad 2a: We state without proof that for any $\alpha>0$:

$$e(-c) + e(c + \alpha) \geqq 1.$$

The proof is completed by application of part 2b, since for any positive number n: $e(-c-n)+e(c+2n)\geqq 1$.

NOTATION. *If F is a σ-complete F-space, if B is the Boolean algebra associated with F, and if Ω is the space defined in terms of B, then denote by Φ the subset of Ω consisting of all those functions which arise from elements of F by the Freudenthal process. We shall denote this relationship*:

$$F \to \Phi \subseteq \Omega.$$

The set Φ may be a proper subset of Ω. For example, let the original F be the space L^p, $p\geqq 1$, on the unit interval. Then the Boolean algebra, B will be the algebra of equivalent classes of measurable sets, reduced modulo sets of measure zero. Let $f(x)$ be a fixed element of L^p $(c\in F)$. Then e_α will be the equivalence class corresponding to

$$\underset{x}{E}[f(x) > \alpha].$$

The space Ω will correspond to the space of all measurable functions, and will be a proper extension of F.

Our problem now is to study the relation between the operations of the space F and those of the space Ω.

THEOREM 2.2. *The space Ω is an extension of the space F. That is, the correspondence between the elements of F and Φ is an isomorphism with respect to ordering,* "sup," "inf," *addition, and multiplication by real numbers.*

Proof. We shall use the following notation: $a, b, c, d \in F$; $f, g, h, k \in \Omega$; $a \rightarrow f$ means that f is the function in Φ corresponding to a in F. $\bigvee_n(a_n)$ is the $\bigvee$ in F. $\bigvee_n(f_n)$ is the $\bigvee$ in Ω. (Similarly with $\bigwedge$, addition, and multiplication by real numbers.)

LEMMA 1. *Ordering is preserved.*

If $c \rightarrow f$, $d \rightarrow g$, then $c \leqq d \rightarrow f \leqq g$. From the Freudenthal integral representation we conclude that $f \leqq g \rightarrow c \leqq d$. Combining these statements with the fact that $c = d \leftrightarrow f = g$, we have

$$c < d \leftrightarrow f < g.$$

LEMMA 2. *Suprema correspond. That is, if $c_n \rightarrow f_n$, and if $\bigvee_n c_n$ exists and is equal to c, then $f = \bigvee_n f_n$ exists and*

$$c \rightarrow f.$$

Since the c_n are bounded above in F, and since order is maintained by the correspondence, the f_n are bounded above in Φ, and therefore in Ω. Therefore $f = \bigvee_n f_n$ exists in Ω. It remains to be proved that

$$\bigvee_{n=1}^{\infty} \bigvee_{m=1}^{\infty} \bigwedge [m(c_n - \alpha)^+, 1] = \bigvee_{m=1}^{\infty} \bigwedge [m(\bigvee_n c_n - \alpha)^+, 1].$$

Now

$$\begin{aligned} \bigwedge [m(\bigvee_n c_n - \alpha)^+, 1] &= \bigwedge [m \bigvee (\bigvee_n c_n - \alpha, 0), 1] \\ &= \bigwedge [m \bigvee_n \bigvee (c_n - \alpha, 0), 1] = \bigwedge [\bigvee_n m(c_n - \alpha)^+, 1]. \end{aligned}$$

By the distributive law the last expression is equal to

$$\bigvee_n \bigwedge [m(c_n - \alpha)^+, 1],$$

and proof is completed by application of the associative law.

LEMMA 3. *Finite infima correspond.*

It is sufficient to prove this for two elements. With a simplification of the notation, we wish to show that

$$\bigwedge \{\bigvee_n \bigwedge (nc^+, 1), \bigvee_n \bigwedge (nd^+, 1)\} = \bigvee_n \bigwedge [n(\bigwedge (c, d))^+, 1].$$

Because of the monotoneity of $\wedge(nc^+, 1)$ and $\wedge(nd^+, 1)$, the left side can be written $\vee_n \wedge \{\wedge(nc^+, 1), \wedge(nd^+, 1)\}$. Application of the distributive law completes the proof.

LEMMA 4. *If* $c \to f$, $-c \to g$, *then* $f+g \geqq 0$.

We wish to prove that for any a

$$1 - \bigvee_n \wedge [n(-a)^+, 1] \leqq \bigwedge_{m=1}^{\infty} \bigvee_n \wedge [n(a + 1/m)^+, 1].$$

The left side is equal to

$$\bigwedge_{n=1}^{\infty} \wedge [n \vee (a + 1/n, 0), 1].$$

For any $m > 0$, and $n \geqq m$:

$$\wedge [n \vee (a + 1/n, 0), 1] \leqq \wedge [n \vee (a + 1/m, 0), 1],$$

and since the first term $\downarrow$ as $n \uparrow$,

$$\bigwedge_{n=1}^{\infty} \wedge [n \vee (a + 1/n, 0), 1] = \bigwedge_{n>m} \wedge [n \vee (a + 1/n, 0), 1]$$

$$\leqq \bigvee_{n=1}^{\infty} \wedge [n(a + 1/m)^+, 1].$$

This is true for every m, and therefore the left side is less than or equal to

$$\bigwedge_{m=1}^{\infty} \bigvee_{n=1}^{\infty} \wedge [n(a + 1/m)^+, 1].$$

LEMMA 5. *If* $a \to f$, $b \to g$, $a+b \to h$, *then* $h \geqq f+g$.

The problem is to show that

$$\bigvee_{\rho} \wedge \left\{ \bigvee_{n=1}^{\infty} \wedge [n(a - \rho)^+, 1], \bigvee_{n=1}^{\infty} \wedge [n(b + \rho - \alpha)^+, 1] \right\}$$

$$\leqq \bigvee_{n=1}^{\infty} \wedge [n(a + b - \alpha)^+, 1].$$

This will be accomplished if we can show that for any a and b

$$\wedge \left\{ \bigvee_{n=1}^{\infty} \wedge (na^+, 1), \bigvee_{n=1}^{\infty} \wedge (nb^+, 1) \right\} \leqq \bigvee_{n=1}^{\infty} \wedge [n(a + b)^+, 1].$$

Because of the monotoneity of $\wedge(na^+, 1)$ and $\wedge(nb^+, 1)$ this relation can be written

$$\bigvee_{n=1}^{\infty} \wedge [(na)^+, (nb)^+, 1] \leqq \bigvee_{n=1}^{\infty} \wedge [(na + nb)^+, 1].$$

This will be proved if we can show that for any a and b

$$\wedge (a^+, b^+, 1) \leqq \wedge [(a + b)^+, 1].$$

This is true since

$$\wedge (a^+, b^+) \leqq (a + b)^+.$$

Proof of this statement is omitted.

LEMMA 6. *The zero elements of F and Ω correspond.*

Proof is omitted.

LEMMA 7. *Negatives of corresponding elements correspond.*

Suppose $c \to f$, $-c \to g$. Then by Lemma 4, $f+g \geqq 0$. But $c+(-c)=0 \to 0$, and by Lemma 4, $0 \geqq f+g$. Therefore $f+g=0$, or $-c \to -f$.

LEMMA 8. *Sums of corresponding elements correspond.*

Let $a \to f$, $b \to g$, $a+b \to h$. Then by Lemma 7,

$$-a \to -f, \; -b \to -g, \; -a + (-b) = -(a+b) \to -h.$$

Therefore by Lemma 5, $h \geqq f+g$, and $-h \geqq (-f)+(-g) = -(f+g)$, so that $h=f+g$.

LEMMA 9. *Infima correspond.*

Suppose $c_n \to f_n$. Then $-c_n \to -f_n$. Consequently

$$\bigvee_n (-c_n) \to \bigvee_n (-f_n)$$

and

$$\bigwedge_n (c_n) = -\bigvee_n (-c_n) \to -\bigvee_n (-f_n) = \bigwedge_n (f_n).$$

LEMMA 10. *Multiplication by real numbers maintains correspondence.*

Because of Lemma 7, and because multiplication by zero has already been considered, we need prove this only for positive numbers. But this is simple routine verification.

This completes the proof of Theorem 2.2.

If we apply the Freudenthal method of extracting a Boolean algebra from a σ-complete F-space to the space Ω, we shall obtain precisely the characteristic functions. Therefore if we construct the space Ω a second time, we obtain no further extension. The space Ω is a maximal extension.

We have already seen that the zero elements of the spaces F and Ω correspond. The same is true of the elements 1. In fact, we can state

THEOREM 2.3. *The correspondence between the elements of the Boolean algebra B and the characteristic functions of Ω coincides with the correspondence between the elements of F and those of Φ.*

Proof is omitted.

Our statement regarding a maximal extension for any σ-complete F-space can now be worded:

THEOREM 2.4. *All spaces with the same Boolean algebra possess a common extension having the same Boolean algebra.*

Freudenthal [1] states without proof that if B is a σ-complete Boolean algebra, then a σ-complete F-space can be constructed having B as its Boolean algebra. The statement is not made in terms of maximal extension, nor is the ring property mentioned, although a space having the properties of Ω was undoubtedly intended. Whether the construction of this space was to be similar to the present one, or more on the order of that of Carathéodory, we do not know.

III. THE SPACE M OF BOUNDED FUNCTIONS

We consider in this chapter the analogue of the space of bounded measurable functions. It is easy to see that if Ω is a σ-complete F-space then the subspace M of bounded functions is also a σ-complete F-space. But it is also a complete normed space. We introduce a norm first for the elements of M^+.

DEFINITION. *If $f \in M^+$, $\|f\| = \text{l.u.b. } \{\alpha\}$ where $f(\alpha) > 0$.*

LEMMA. 1. $\|0\| = 0$. *If $f > 0$, $\|f\| > 0$.* 2. *If $0 \leq f \leq g$, $\|f\| \leq \|g\|$.* 3. *If $f \geq 0$, $\alpha \geq 0$, then $\|\alpha f\| = \alpha\|f\|$.* 4. *If f, $g \geq 0$, then $\|f+g\| \leq \|f\| + \|g\|$.*

Proof is omitted.

DEFINITION. *For arbitrary $f \in M$, $\|f\| = \| |f| \|$.*

THEOREM 3.1. *The space M is normed.*

Proof. 1. $\|f\| \geq 0$. $\|f\| = 0 \leftrightarrow f = 0$. 2. $\|f+g\| \leq \|f\| + \|g\|$. The first is trivial and the second is true by the lemma above and because $|f+g| \leq |f| + |g|$.

THEOREM 3.2. *The space M is complete.*

Proof. We first establish a lemma, using the notation $f_n \to^o f$ to represent (o)-convergence (cf. Birkhoff [8, chap. 2, §36]).

LEMMA. *If f_n is a sequence of elements of M^+ such that* (1) $f_n \to^o 0$, (2) $\|f_n - f_m\| \to 0$, *then*

$$\|f_n\| \to 0.$$

Proof of Lemma. By the assumption (2) above, given any $\beta>0$, we can find an N such that for $n, m \geqq N$,

$$(f_n - f_m)(\beta) = \bigvee_\rho \bigwedge \left[f_n(\rho), 1 - \bigwedge_{p=1}^{\infty} f_m(-\beta+\rho-1/p) \right] = 0.$$

Because of the monotoneity of $f_m(-\beta+\rho-1/p)$ as a function of p,

$$\bigwedge_{p=1}^{\infty} f_m(-\beta+\rho-1/p) = \bigwedge_{p>1/\beta}^{\infty} f_m(-\beta+\rho-1/p) \leqq f_m(-2\beta+\rho),$$

and therefore for any ρ

$$\bigwedge [f_n(\rho), 1 - f_m(-2\beta+\rho)] = 0.$$

In particular we may choose $\rho=3\beta$, in which case we can state that for all $n, m \geqq N$, $\bigwedge[f_n(3\beta), 1-f_m(\beta)]=0$. Therefore

$$\bigvee_{m>N}^{\infty} \bigwedge [f_n(3\beta), 1 - f_m(\beta)] = \bigwedge \left[f_n(3\beta), 1 - \bigwedge_{m>N}^{\infty} f_m(\beta) \right] = 0.$$

But since $\bigwedge_{m>N}^{\infty} f_m(\beta)=0$ by condition (1), $f_n(3\beta)=0$ for all $n \geqq N$, and therefore $\|f_n\| \to 0$.

Before completing the proof of Theorem 3.2 we remark that, for an arbitrary sequence f_n of elements of M,

$$\|f_n\| \to 0 \rightarrow f_n \to^o 0$$

but the reverse implication is not true.

Assuming we have a sequence f_n of elements of M such that

$$\|f_n - f_m\| \to 0,$$

we wish to find an f in M such that

$$\|f - f_n\| \to 0.$$

Following the remark above, we can state that

$$f_n - f_m \to^o 0,$$

and therefore, as proved by Kantorovitch [3], there exists an element f of M such that

$$f - f_n \to^o 0.$$

We shall prove that

$$\|f - f_n\| \to 0.$$

Let $F_n = f_n - f$. Then $|F_n| \to^o 0$. And, since $|F_n - F_m| \geqq ||F_n| - |F_m||$, we have

(1) $|F_n| \to^o 0$,
(2) $\| |F_n| - |F_m| \| \to 0$.

Therefore, by the preceding lemma,

$$\| |F_n| \| = \|F_n\| = \|f_n - f\| \to 0.$$

This completes the proof of Theorem 3.2.

We have seen that convergence according to ordering does not imply convergence according to the norm. Furthermore the space M need not be regular (cf. Kantorovitch [3]). For example, the space of bounded measurable functions on the unit interval is not regular.

IV. The Λ^p spaces

In this chapter we shall be concerned with a Boolean algebra B with a completely additive bounded measure.

Definition, *A Boolean algebra of type B_μ is a σ-complete Boolean algebra B with a real-valued (measure) function $\mu(a)$ defined for every element a of B and such that*

1. $0 \leqq \mu(a) \leqq 1$ *for every a of B,*
2. $\mu(a) = 0 \leftrightarrow a = 0$,
3. $\mu(1) = 1$,
4. *if* $\wedge(a, b) = 0$, $\mu \vee(a, b) = \mu(a) + \mu(b)$,
5. *if* $a_n \downarrow 0$, $\mu a_n \downarrow 0$.

Remarks. 1. If $a_n \downarrow a, \mu a_n \downarrow \mu a$; if $a_n \uparrow a, \mu a_n \uparrow \mu a$. 2. If $(a_1, a_2, \cdots)$ is any non-empty countable set of elements of B, then

$$\mu \vee (a_1, a_2, \cdots) \leqq \mu(a_1) + \mu(a_2) + \cdots.$$

If the elements are pairwise disjoint we have equality.

Theorem 4.1. *A Boolean algebra B of type B_μ is complete. In fact, if F is any subset of B, then there is a countable subset $F' \subseteq F$ such that $\vee(F') = \vee(F)$.*

Proof is given by Wecken [7].

In this section we construct spaces similar to the Lebesgue L^p spaces and belonging to the class of what Kantorovitch [3] calls spaces of type B_2, which we shall call normed K-spaces. They are defined:

Definition. *A normed K-space is a σ-complete K-space K with a norm $\|f\|$ satisfying the following conditions for the elements of K^+:*

1. $\|0\| = 0$; $\|f\| > 0$ *if* $f > 0$;
2. $\|f + g\| \leqq \|f\| + \|g\|$;
3. $\|\alpha f\| = \alpha \|f\|$ *where* $\alpha > 0$;
4. *if* $f < g$, *then* $\|f\| < \|g\|$;
5. *if* $f_n \downarrow 0$, *then* $\|f_n\| \downarrow 0$; *if* $f_n \uparrow \infty$, *then* $\|f_n\| \uparrow \infty$.

For arbitrary elements, the norm is given by

$$\|f\| = \| \, |f| \, \|.$$

DEFINITION. *A normed F-space is a σ-complete F-space, where the element* 1 *has norm equal to* 1.

DEFINITION. *A space of type* $K^p(F^p)$, $p \geqq 1$, *is a σ-complete K-space (F-space) such that* $\wedge(f, g) = 0 \rightarrow \|f+g\|^p = \|f\|^p + \|g\|^p$.

We shall discover a close analogy between spaces of type F^p and the Lebesgue L^p spaces. A similar problem, from the axiomatic viewpoint, has been studied by Bohnenblust.

Let us reverse the direction of investigation for a moment and consider the Boolean algebra of a space of type F^p, $p \geqq 1$, when the latter is given. Define the function on the elements of B:

$$\mu(a) = \|a\|^p.$$

Then it is easily seen that the Boolean algebra is of type B_μ.

Returning to the main problem, let B be a Boolean algebra of type B_μ, and let Ω be the space of functions on B. Let p be any fixed real number $\geqq 1$, and consider the elements of Ω^+.

DEFINITION. *If s is the simple function*

$$s = \alpha_1 a_1 + \cdots + \alpha_n a_n,$$

where $\alpha_i \geqq 0$, $i = 1, 2, \cdots, n$, and $\wedge(a_i, a_j) = 0$, $i \neq j$, *then we define its integral*:

$$\int s d\mu = \alpha_1 \mu a_1 + \cdots + \alpha_n \mu a_n,$$

and its norm:

$$\|s\| = \left[\int s^p d\mu \right]^{1/p} = \left[\alpha_1^p \mu a_1 + \cdots + \alpha_n^p \mu a_n\right]^{1/p}.$$

For an arbitrary $f \in \Omega^+$ we define the integral:

$$\int f d\mu = \sup \left\{ \int s d\mu \right\} \quad \text{for all simple functions } s, \qquad 0 \leqq s \leqq f.$$

If we denote by M^p the set of all $f \in \Omega^+$ such that $\int f^p d\mu < \infty$, we can define the norm of f, for any $f \in M^p$:

$$\|f\| = \left[\int f^p d\mu \right]^{1/p}.$$

In order to prove that this norm satisfies the necessary conditions, we establish a number of lemmas:

LEMMA 1. *If* $g \in M^p$, $0 \leqq f \leqq g$, *then* $f \in M^p$ *and* $\|f\| \leqq \|g\|$.

Proof is trivial. In particular, every bounded function of Ω^+ belongs to M^p for every $p \geqq 1$.

LEMMA 2. *If* s *and* t *are simple functions*, $\geqq 0$, *then*

$$\int (s+t)d\mu = \int s d\mu + \int t d\mu,$$

$$\|s+t\| \leqq \|s\| + \|t\|.$$

Proof rests on the fact that s and t, and therefore $s+t$, can be represented as linear combinations, with non-negative coefficients, of the same set of pairwise disjoint characteristic functions. The second relation follows from an application of Minkowski's inequality.

In order to extend these relations to arbitrary functions in M^p, we prove

LEMMA 3. *If* $\{s_n\}$ *is an increasing sequence of positive simple functions whose limit is the function* f, *then*

$$\int s_n d\mu \uparrow \int f d\mu,$$

whether the limit be finite or infinite.

Proof. If s is any simple function such that $0 \leqq s \leqq f$, and if $\alpha > 0$, we wish to find an n such that

$$\int s_n d\mu > \int s d\mu - \alpha.$$

The method of obtaining n is straightforward.

From this lemma follows immediately:

LEMMA 4. *If* $f \in M^p$, *and if* $0 \leqq s_n \uparrow f$, *where* s_n *are simple functions, then* $\|s_n\| \uparrow \|f\|$. *Also*

$$\|f\| = \sup \{\|s\|\}$$

for all simple functions s *such that* $0 \leqq s \leqq f$.

In the proof of Theorem 1.12 an increasing sequence of simple functions was given having as limit a given function of Ω^+. From the form of those simple functions and from Lemma 3 we have

LEMMA 5. *If* $f \in \Omega^+$, *then* $\int f d\mu$, *finite or infinite, is equal to the least upper bound of sums of the form:*

$$\alpha_1 \mu f(\alpha_1) + (\alpha_2 - \alpha_1)\mu f(\alpha_2) + \cdots + (\alpha_n - \alpha_{n-1}) u f(\alpha_n),$$

where $0 \leqq \alpha_1 \leqq \cdots \leqq \alpha_n$.

Similarly we have

LEMMA 6. *If $f \in M^p$, then $\|f\|^p$ is equal to the least upper bound of sums of the form:*

$$\alpha_1^p \mu f(\alpha_1) + \cdots + (\alpha_n^p - \alpha_{n-1}^p)\mu f(\alpha_n),$$

where $0 \leqq \alpha_1 \leqq \cdots \leqq \alpha_n$.

THEOREM 4.2. *If $f, g \in M^p$, $p \geqq 1$, then $f+g \in M^p$ and*

$$\|f + g\| \leqq \|f\| + \|g\|.$$

Proof. Let $0 \leqq s_n \uparrow f$, $0 \leqq t_n \uparrow g$. Then $0 \leqq s_n + t_n \uparrow f+g$ and therefore $\|s_n + t_n\| \uparrow \|f+g\|$. But $\|s_n + t_n\| \leqq \|s_n\| + \|t_n\| \leqq \|f\| + \|g\|$ for every n. The inequality follows.

THEOREM 4.3. *If $f, g \in M^1$, then $f+g \in M^1$ and $\|f+g\| = \|f\| + \|g\|$.*

Proof. Let $0 \leqq s_n \uparrow f$, $0 \leqq t_n \uparrow g$; and given $\alpha > 0$, choose n so that $\int f d\mu < \int s_n d\mu + \alpha$, $\int g d\mu < \int t_n d\mu + \alpha$. Then

$$\int (f + g) d\mu \geqq \int (s_n + t_n) d\mu > \int f d\mu + \int g d\mu - 2\alpha.$$

The implied inequality combined with that of the previous theorem, gives the desired equality.

LEMMA. *If $\wedge(a, b) = 0$, then*

$$f_a \in M^p, f_b \in M^p \longleftrightarrow f_{\vee(a,b)} \in M^p$$

and

$$\|f_{\vee(a,b)}\|^p = \|f_a\|^p + \|f_b\|^p.$$

Proof. We may assume $p = 1$, and proof follows from Theorem 4.3.

THEOREM 4.4. *If $\wedge(f, g) = 0$, then $f, g \in M^p$ if and only if $f+g \in M^p$, and $\|f+g\|^p = \|f\|^p + \|g\|^p$.*

Proof. In the preceding lemma set $a = f(0)$, $b = g(0)$.

DEFINITION. *Let Λ^p be the subspace of Ω consisting of all f such that $|f| \in M^p$. If $f \in \Lambda^p$, we define the norm of f:*

$$\|f\| = \| \, |f| \, \|.$$

If $f \in \Lambda^1$, we define the integral of f:

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

Remarks. 1. $M^p = [\Lambda^p]^+$, 2. if $1 \leqq p \leqq q$, $\Lambda^q \subseteq \Lambda^p$.

THEOREM 4.5. *If* f, $g \in \Lambda^1$, *then* $f+g \in \Lambda^1$, *and*

$$\int (f+g)d\mu = \int fd\mu + \int gd\mu.$$

Proof. Since $|f+g| \leqq |f| + |g|$, $f+g \in \Lambda^1$. Applying the additivity of the integral for elements of $[\Lambda^1]^+$ to the known equality $(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+$, and rearranging terms, we have the desired relation:

$$\int (f+g)^+ d\mu - \int (f+g)^- d\mu = \int f^+ d\mu - \int f^- d\mu + \int g^+ d\mu - \int g^- d\mu.$$

We now state the fundamental theorem of this section:

THEOREM 4.6. Λ^p *is a space of type* F^p, $p \geqq 1$.

Proof. 1. If f, $g \in \Lambda^p$, then $f+g \in \Lambda^p$, since $|f+g| \leqq |f| + |g|$; 2. if $f \in \Lambda^p$, α real, then $\alpha f \in \Lambda^p$, since $|\alpha f| = |\alpha|\,|f|$; 3. if $f \in \Lambda^p$, then $f^+ \in \Lambda^p$, since $f^+ \leqq |f|$; 4. if $f \leqq g \leqq h$, f, $h \in \Lambda^p$, then $g \in \Lambda^p$, since $|g| \leqq h^+ + f^-$; 5. $\|1\| = 1$. Therefore Λ^p is a σ-complete F-space. We now verify the five conditions of the definition of a normed K-space.

Ad 1, 2, 3: Already proved, or trivial.

Ad 4: If $0 \leqq f < g$, then $0 \leqq f^p < g^p$, and therefore

$$\|f\|^p = \int f^p < \int g^p = \|g\|^p.$$

Ad 5: We state without proof the lemma:

LEMMA. *If* $f_n \geqq 0$, $g \geqq 0$ *are elements of a vector lattice, and if* $f_n \downarrow$, *then* $f_n - \wedge(f_n, g) \downarrow$.

Assume that $f_n \downarrow 0$, and that α is any number >0. We can find an N such that $\int [f_1 - \wedge(f_1, N)]d\mu < \alpha$. Then for every n $\int [f_n - \wedge(f_n, N)]d\mu < \alpha$. In order to make $\int f_n d\mu < 2\alpha$, we wish to find an n such that $\int \wedge(f_n, N)d\mu < \alpha$. An n which will accomplish this will be any one such that

$$\mu[f_n(\alpha/2)] < \alpha/2N.$$

Now assume that $0 \leqq f_n \uparrow \infty$, and that $\|f_n\| \leqq M$ for every n. We shall obtain a contradiction. Define

$$f(\xi) = \bigvee_n f_n(\xi).$$

Then $f(\xi)$ must certainly belong to Ω, or we easily obtain a contradiction. Now form the sum, for $0 < \beta_1 < \cdots < \beta_n$:

$$\beta_1 \mu f(\beta_1) + \cdots + (\beta_n - \beta_{n-1})\mu f(\beta_n).$$

Given $\alpha > 0$, choose $k_1, \cdots, k_n$ such that

$$\mu[f_{k_1}(\beta_1)] > \mu[f(\beta_1)] - \alpha/\beta_n,$$
$$\cdots\cdots\cdots\cdots\cdots,$$
$$\mu[f_{k_n}(\beta_n)] > \mu[f(\beta_n)] - \alpha/\beta_n.$$

Letting $N = \max(k_1, \cdots, k_n)$, we have

$$\beta_1\mu f(\beta_1) + \cdots + (\beta_n - \beta_{n-1})\mu f(\beta_n)$$

is less than

$$\beta_1[\mu f_N(\beta_1) + \alpha/\beta_n] + \cdots + (\beta_n - \beta_{n-1})[\mu f_N(\beta_n) + \alpha/\beta_n]$$
$$\leqq \beta_1\mu f_N(\beta_1) + \cdots + (\beta_n - \beta_{n-1})\mu f_N(\beta_n) + \frac{\alpha}{\beta_n}\beta_n \leqq M + \alpha.$$

Therefore $f \in \Lambda^p$. (Contradiction.) This completes the proof of Theorem 4.6.

Remark. As proved by Kantorovitch [3], a normed K-space is regular, and convergence according to the norm is the same as star-convergence. This statement is therefore true for the Λ^p spaces. Let us observe also that the simple functions are dense, in the sense of the norm, in Λ^p for every $p \geqq 1$.

THEOREM 4.7. *Every space of type F^p, $p \geqq 1$, is isomorphic to the space Λ^p, where the basic Boolean algebra of Λ^p is the Boolean algebra of F^p.*

Proof. We have seen that the Boolean algebra of F^p is of type B_μ, and that therefore the corresponding spaces Ω and Λ^p can be constructed. As proved in the second chapter, the space F^p is isomorphic to a subspace of Ω. This subspace is precisely the space Λ^p. The isomorphism preserves norm. Details of the proof are omitted.

V. Absolute continuity

In this chapter we shall study the space AC of additive absolutely continuous real-valued functions defined on a Boolean algebra of type B_μ, and set up an isomorphism between AC and Λ^1.

DEFINITION. *A real-valued function, $\xi(a)$, defined on the elements of a Boolean algebra B, of type B_μ, is absolutely continuous if, given any $\alpha > 0$, there exists a $\beta > 0$ such that*

$$\mu(a) < \beta \rightarrow |\xi(a)| < \alpha.$$

DEFINITION. *The space AC is the space of all additive absolutely continuous real-valued functions, $\xi(a)$, defined on the elements of a Boolean algebra B of type B_μ, with the usual definitions of addition, multiplication by real numbers, and ordering.*

We shall show that AC is a space of type F^1 by obtaining an isomorphism

between AC and Λ^1. First we set up a one-to-one correspondence between AC^+ and $[\Lambda^1]^+$.

DEFINITION. *We define the integral of a function,* $f\in\Lambda^1$, *over an element of the Boolean algebra* B:

$$\int_a f d\mu \equiv \int f_a d\mu.$$

DEFINITION: *The transformation* X *is defined by the equation*

$$\xi(a) \equiv X[f(\xi)]$$

where $f\in[\Lambda^1]^+$ *and* $\xi(a)\equiv\int_a f d\mu$. *The transformation* X *is an integration operation. We say that* $\xi(a)$ *is the indefinite integral of* $f(\xi)$.

THEOREM 5.1. *If* $f\in[\Lambda^1]^+$, *then* $X(f)\in AC^+$.

Proof. Positiveness and finite additivity are easily verified. Let $\alpha>0$, and let $f=f_1+f_2$, where f_1 and f_2 belong to $[\Lambda^1]^+$, f_1 is bounded by N, and $\int f_2 d\mu<\alpha$. Then

$$\int_a f d\mu < \int_a f_1 d\mu + \alpha.$$

Now, if $\mu(a)<N/\alpha$, we have $\xi(a)=\int_a f d\mu<2\alpha$.

THEOREM 5.2. *Let* $\xi(a)\in AC^+$ *and let* $\alpha\geqq 0$. *Form the set* S_α *consisting of the zero of* B *and all* a *of* B *such that*

$$0 < b \leqq a \rightarrow \xi(b) > \alpha\cdot\mu(b).$$

Then S_α *is a principal ideal. That is, if*

$$a \leqq \lambda(\alpha) = \bigvee(S_\alpha)$$

then $a\in S_\alpha$.

Proof. 1. If $a_1\in S_\alpha$, $a_2\leqq a_1$, then $a_2\in S_\alpha$. 2. If $a_1, a_2\in S_\alpha$, then $\bigvee(a_1, a_2)\in S_\alpha$. 3. If $a_1, a_2, \cdots \in S_\alpha$, then $\bigvee(a_1, a_2, \cdots)\in S_\alpha$. 4. If $Q\subseteq S_\alpha$, then $\bigvee(Q)\in S_\alpha$.

Ad 2: We may assume a_1 and a_2 to be disjoint. Then if $0<b\leqq\bigvee(a_1, a_2)$, write $b=\bigvee(b_1, b_2)$, where $b_1=\bigwedge(b, a_1)$, $b_2=\bigwedge(b, a_2)$. Then $\xi(b_1)\geqq\alpha\mu(b_1)$, $\xi(b_2)\geqq\alpha\mu(b_2)$, where the strict $>$ relation must hold in at least one case since not both b_1 and b_2 can be zero. Therefore $\xi(b)>\alpha\mu(b)$.

Ad 4: In a Boolean algebra of type B_μ any "sup" can be reduced to a countable one.

DEFINITION. *The transformation* Y *is defined by the equation*

$$f(\xi) \equiv Y[\xi(a)]$$

where $\xi(\mathbf{a}) \in AC^+$ *and*

$$f(\xi) \equiv \begin{cases} 1, & \xi < 0, \\ \bigvee (S_\xi), & \xi \geqq 0. \end{cases}$$

The transformation Y is a differentiation operation. We say that $f(\xi)$ is the derivative of $\xi(\mathbf{a})$.

THEOREM 5.3. *If $f \in [\Lambda^1]^+$, then $Y[X(f)] = f$. That is, the derivative of the integral of a function $f \in [\Lambda^1]^+$ is f.*

Proof. Given any $\alpha \geqq 0$, we wish to show:

1. $f(\alpha) \in S_\alpha$.
2. $a \in S_\alpha \rightarrow a \leqq f(\alpha)$.

Ad 2: Assume there exists an $a > 0$ such that the inequality $a \leqq f(\alpha)$ does not hold, $0 < b \leqq a \rightarrow \int_b f d\mu > \alpha\mu(b)$. But if we take $b = a - \bigwedge[a, f(\alpha)]$, then $0 < b \leqq a$, and therefore $\int_b f d\mu > \alpha\mu(b)$. This is absurd since $f_b(\alpha) = \bigwedge[b, f(\alpha)] = 0$. The case $S_\alpha = 0$ is trivial.

THEOREM 5.4. *Let $\xi(\mathbf{a}) \in AC^+$, $\alpha \geqq 0$, and let S_α be the principal ideal defined above. If $a = \bigvee(S_\alpha)$, $\bigwedge(a, c) = 0$, then $\xi(c) \leqq \alpha\mu(c)$.*

Proof. Suppose $0 < c$, $\bigwedge(a, c) = 0$, $\xi(c) > \alpha\mu(c)$. Let $\beta_n \downarrow 0$.

Let m_1 be the least upper bound of $\mu(d)$ for all $d < c$ such that $\xi(d) \leqq \alpha\mu(d)$. Then $m_1 > 0$. Find c_1 such that

$$0 < c_1 < c, \qquad \xi(c_1) \leqq \alpha\mu(c_1), \qquad \mu(c_1) > m_1 - \beta_1.$$

In general let $m_n > 0$ be the least upper bound of $\mu(d)$ for all d such that $d < c - c_1 - \cdots - c_{n-1}$; $\xi(d) \leqq \alpha\mu(d)$. Pick c_n such that

$$0 < c_n < c - c_1 - \cdots - c_{n-1}, \qquad \xi(c_n) \leqq \alpha\mu(c_n), \qquad \mu(c_n) > m_n - \beta_n.$$

Regarding the measures, we must have $m_n \leqq \beta_{n-1}$, $n \geqq 2$. For suppose there is an n such that $m_n > \beta_{n-1}$. Then find an element $b < c - c_1 - \cdots - c_{n-1}$ such that $\mu(b) > \beta_{n-1}$ and $\xi(b) \leqq \alpha\mu(b)$. Then $\mu(c_{n-1} + b) > m_{n-1} - \beta_{n-1} + \beta_{n-1} = m_{n-1}$. But since $c_{n-1} + b < c - c_1 - \cdots - c_{n-2}$, and $\xi(c_{n-1} + b) \leqq \alpha\mu(c_{n-1} + b)$, we know that $\mu(c_{n-1} + b) \leqq m_{n-1}$. (Contradiction.)

Form $c' = \bigvee_{n=1}^{\infty} c_n \leqq c$. Then by absolute continuity, $\xi(c') \leqq \alpha\mu(c')$, and therefore $c' < c$ and $\xi(c - c') > \alpha\mu(c - c')$. Therefore there exists a d such that $0 < d < c - c'$, $\xi(d) \leqq \alpha\mu(d)$.

Now choose an n such that $\beta_{n-1} < \mu(d)$. Then $m_n < \mu(d)$. Therefore, since $d < c - c_1 - \cdots - c_{n-1}$, $\mu(d) \leqq m_n$. (Contradiction.)

THEOREM 5.5. *If $\xi(\mathbf{a}) \in AC^+$, then $f(\xi) \equiv Y[\xi(\mathbf{a})] \in [\Lambda^1]^+$.*

Proof. We shall first verify that $f \in \Omega^+$. It is sufficient to verify parts 1, 2b, and 3 in the definition of Ω.

Ad 1: If $0 \leqq \alpha < \beta$, $S_\beta \subseteq S_\alpha$, $f(\beta) \leqq f(\alpha)$.

Ad 2b: Suppose that for every $n = 1, 2, \cdots$

$$\bigvee (S_n) \geqq d > 0.$$

Then $d \in S_n$ for every n, and therefore $\xi(d) > n\mu(d)$ for every n. But this is absurd.

Ad 3: We wish to prove that for any $\alpha \geqq 0$

$$\bigvee_{n=1}^{\infty} \bigvee (S_{\alpha+1/n}) < \bigvee (S_\alpha)$$

leads to a contradiction. Obviously the right side is >0. Let $b_n = \bigvee(S_\alpha) - \bigvee(S_{\alpha+1/n})$. Then $b_n \downarrow b > 0$. Therefore $\xi(b) > \alpha\mu(b)$. Pick an M such that $\xi(b) > (\alpha + 1/M)\mu(b)$. Then there exists an $N > M$ such that $\xi(b) > (\alpha + 1/M)\mu(b_N)$, so that $\xi(b_N) > (\alpha + 1/N)\mu(b_N)$. Since $\bigwedge [b_N, \bigvee(S_{\alpha+1/N})] = 0$, the necessary contradiction is provided by Theorem 5.4.

In verifying that $f(\xi) \in [\Lambda^1]^+$, we shall prove more. Let a be an arbitrary positive element of B, and, for an arbitrary set of real numbers, $\alpha_1 < \alpha_2 < \cdots < \alpha_n$, form

$$\alpha_1\mu \wedge [a, f(\alpha_1)] + \cdots + \alpha_n - (\alpha_{n-1})\mu \wedge [a, f(\alpha_n)].$$

Defining $b_i = \bigwedge [a, f(\alpha_i)]$, we can write this sum:

$$\alpha_1\mu(b_1 - b_2) + \alpha_2\mu(b_2 - b_3) + \cdots + \alpha_n\mu(b_n).$$

Since $b_i - b_{i+1} \leqq b_i \leqq f(\alpha_i)$, we have $\xi(b_i - b_{i+1}) \geqq \alpha_i\mu(b_i - b_{i+1})$, where the equality holds only if $b_i = b_{i+1}$. Therefore the previous sum is less than or equal to

$$\xi(b_1 - b_2) + \xi(b_2 - b_3) + \cdots + \xi(b_{n-1} - b_n) + \xi(b_n) = \xi(b_1) \leqq \xi(a).$$

Therefore $\int_a f d\mu = \int f_a d\mu \leqq \xi(a)$. The proof of the theorem is completed when we set $a = 1$.

The last inequality above implies the

LEMMA. *If $\xi(\mathbf{a}) \in AC^+$, then $X[Y(\xi(\mathbf{a}))] \leqq \xi(\mathbf{a})$.*

The inequality can be replaced by an equality:

THEOREM 5.6. *If $\xi(\mathbf{a}) \in AC^+$, then $X[Y(\xi(\mathbf{a}))] = \xi(\mathbf{a})$. That is, every additive absolutely continuous function $\xi(\mathbf{a}) \geqq 0$ is an integral. It is equal to the integral of its derivative.*

Proof. We shall establish this theorem by means of lemmas:

LEMMA 1. *If $h = f_a$, and if $\alpha \geqq 0$, then $h(\alpha) = \bigvee(S_{\alpha,a})$, where $S_{\alpha,a}$ is the set consisting of zero and all positive elements c of B such that $c \leqq a$, and $0 < b \leqq c \rightarrow \xi(b) > \alpha\mu(b)$.*

Proof is trivial.

LEMMA 2. *If* $\alpha \geqq 0, \beta \geqq 0$, *then*

$$\alpha\mu[h(\alpha) - h(\alpha+\beta)] \leqq \xi[h(\alpha) - h(\alpha+\beta)] \leqq (\alpha+\beta)\mu[h(\alpha) - h(\alpha+\beta)].$$

Proof. The first inequality follows from the fact that $h(\alpha)-h(\alpha+\beta) \leqq h(\alpha)$. The second inequality is obvious if $h(\alpha+\beta)=h(\alpha)$. Assume $h(\alpha+\beta)<h(\alpha)$. Then $0<h(\alpha)-h(\alpha+\beta)\leqq a$, $\wedge[h(\alpha)-h(\alpha+\beta), h(a+\beta)]=0$, and the inequality is a consequence of Theorem 5.4.

This result can be restated as

LEMMA 3. *If* $\rho>0$, $\alpha\geqq 0$, *then for* $0\leqq\beta\leqq\rho$,

$$0 \leqq \xi[h(\alpha) - h(\alpha+\beta)] - \alpha\mu[h(\alpha) - h(\alpha+\beta)] \leqq \rho\mu[h(\alpha) - h(\alpha+\beta)].$$

LEMMA 4. *Let* $C=\vee(S_a)$, $A=\vee(S_{\alpha,a})=\wedge(a, C)$. *If* $\wedge(A, b)=0$, $b\leqq a$, then $\xi(b)\leqq\alpha\mu(b)$.

Proof. Since $\wedge(A, b)=\wedge(a, b, C)=\wedge(b, C)=0$, we can apply Theorem 5.4.

LEMMA 5. *Given* $\rho<0$, *we can find* $\alpha_1<\alpha_2<\cdots<\alpha_n$ *such that*

$$\xi(a) - [\alpha_1\mu h(\alpha_1) + \cdots + (\alpha_n - \alpha_{n-1})\mu h(\alpha_n)] < 3\rho.$$

Proof. By Lemma 4, and since $h(0)=\vee(S_{0,a})$, we have $\xi[a-h(0)]=0$, and therefore we can pick α_1 such that

$$\xi(a) - \xi[h(\alpha_1)] < \rho.$$

Choose $N>\alpha_1$ such that $\xi[h(N)]<\rho$, and finally choose α_1 (as above) $<\alpha_2<\cdots<\alpha_n=N$ such that $\alpha_i-\alpha_{i-1}<\rho$, $i=2, \cdots, n$. Then

$$\begin{aligned}\xi(a) - [\alpha_1\mu(h(\alpha_1) - h(\alpha_2)) + \cdots + \alpha_n\mu h(\alpha_n)]&\\ = \xi[a - h(\alpha_1)] + \xi[h(\alpha_1) - h(\alpha_2)] + \cdots + \xi[h(\alpha_n)]&\\ - \alpha_1\mu[h(\alpha_1) - h(\alpha_2)] - \cdots - \alpha_n\mu h(\alpha_n)&\\ \leqq \rho + \rho[\mu(h(\alpha_1) - h(\alpha_2)) + \cdots + \mu(h(\alpha_{n-1}) - h(\alpha_n))] + \rho&\\ = 2\rho + \rho[\mu(h(\alpha_1) - h(\alpha_n))] \leqq 3\rho.&\end{aligned}$$

This completes the proof of Theorem 5.6.

We shall now consider arbitrary elements of AC.

THEOREM 5.7. *Let* $\xi(\mathbf{a})\in AC$, *and let* S *be the set consisting of zero and all positive elements* a *of* B *such that*

$$0 < b \leqq a \rightarrow \xi(b) > 0.$$

Then S *is a principal ideal.*

Proof. Proof is similar to that of Theorem 5.2.

THEOREM 5.8. *If* $A=\vee(S)$, $\wedge(c, A)=0$, *then* $\xi(c)\leqq 0$.

Proof is similar to that of Theorem 5.4.

As before, we can define S_a to be the set consisting of all elements of the form $\wedge(a, d)$ where $d \in S$. Then S_a is also a principal ideal. We can also establish the

LEMMA. *Let* $C = \vee(S)$, $A = \vee(S_a) = \wedge(a, C)$. *If* $\wedge(A, b) = 0$, *and* $b \leqq a$, *then* $\xi(b) \leqq 0$.

DEFINITION. *Given any* $\xi(\mathbf{a}) \in AC$ *and any* $a \in B$, *let* $a^+ = \vee(S_a)$. *(Then* a^+ *is a function of* $\xi(\mathbf{a})$ *and* a.) *Define* $\xi^+(\mathbf{a})$, $\xi^-(\mathbf{a})$:

$$\xi^+(a) \equiv \xi(a^+), \qquad \xi^-(a) \equiv -\xi(a - a^+).$$

THEOREM 5.9. *If* $\xi(\mathbf{a}) \in AC$, *then* $\xi^+(\mathbf{a})$ *and* $\xi^-(\mathbf{a}) \in AC$ *and* $\xi^+(\mathbf{a}) = \vee[\xi(\mathbf{a}), 0(\mathbf{a})]$, $\xi^-(\mathbf{a}) = \vee[-\xi(\mathbf{a}), 0(\mathbf{a})]$.

Proof is omitted.

The space AC is therefore a vector lattice.

We proceed now to the establishment of the one-to-one correspondence between the elements of AC and those of Λ^1. We extend the definitions of X and Y:

DEFINITION. *Given any element* $f \in \Lambda^1$, *we define*

$$\xi(\mathbf{a}) \equiv X(f) \equiv X(f^+) - X(f^-) \equiv \xi^1(\mathbf{a}) - \xi^2(\mathbf{a}).$$

Given any element $\xi(\mathbf{a}) \in AC$, *we define*

$$f(\xi) \equiv Y[\xi(\mathbf{a})] \equiv Y[\xi^+(\mathbf{a})] - Y[\xi^-(\mathbf{a})] \equiv f^1(\xi) - f^2(\xi).$$

To prove that the transformations X and Y are inverse, we must show that $\xi^1(\mathbf{a}) = \xi^+(\mathbf{a})$, $\xi^2(\mathbf{a}) = \xi^-(\mathbf{a})$, $f^1(\xi) = f^+(\xi)$, and $f^2(\xi) = f^-(\xi)$; in other words, that

$$\wedge\,[\xi^1(\mathbf{a}), \xi^2(\mathbf{a})] = \wedge\,[f^1(\xi), f^2(\xi)] = 0.$$

Proof is omitted.

We have therefore a one-to-one correspondence between the elements of AC and those of Λ^1, and $X^{-1} = Y$, $Y^{-1} = X$.

The transformation $\xi(\mathbf{a}) = X(f)$ can now be represented in general in the form:

$$\xi(a) = \int_a f d\mu.$$

THEOREM 5.10. *The transformations* X *and* Y *preserve addition, ordering, and multiplication by real numbers. That is,* X *and* Y *are additive, homogeneous, and positive*:

$$X(\alpha f + \beta g) = \alpha X f + \beta X g, \qquad Y(\alpha\xi + \beta\rho) = \alpha Y\xi + \beta Y\rho,$$

$$f > 0 \rightarrow X(f) > 0,$$
$$\xi(a) > 0 \rightarrow Y[\xi(a)] > 0.$$

Proof. We need prove only: 1. $\int_a(\alpha f+\beta g)d\mu=\alpha\int_a fd\mu+\beta\int_a gd\mu$. 2. $f\geqq 0 \rightarrow \int_a fd\mu \geqq 0$, for every a. 3. $f>0\rightarrow\int_a fd\mu>0$, for some a. 4. f not $\geqq 0\rightarrow\int_a fd\mu<0$, for some a. Details of the proof are omitted.

This completes the proof that AC is a space of type F^1, since it is isomorphic to Λ^1. Conversely, every space of type F^1 is isomorphic to a space of absolutely continuous functions.

Remark. If a is an element of B, then the function of AC which corresponds to the function a of Λ^1 is the function

$$\xi(b) = \mu[\wedge (a, b)].$$

VI. Concluding remarks

1. **Regularity.** Since a given Boolean algebra B determines and is determined by the corresponding space Ω, we may wish to know the conditions on B corresponding to the regularity conditions on Ω. A set of necessary and sufficient conditions is contained in the

DEFINITION. *A regular Boolean algebra is a σ-complete Boolean algebra B which satisfies the further conditions*:

1. *If F is any subset of B, then $\vee(F)$ exists, and there is a countable subset $F'\subseteq F$ such that*

$$\vee (F') = \vee (F).$$

2. *If $a_n \downarrow 0$ as $n \uparrow \infty$, $a_n^k \downarrow a_n$ as $k \uparrow \infty$, then there exists a subsequence, $a_n^{k_n}$, such that*

$$\bigwedge_{n=1}^{\infty} \bigvee_{m=n}^{\infty} a_m^{k_m} = 0.$$

3. *If $a_m^k(n) \downarrow 0$ as $k \uparrow \infty$, for each m, n and $a_m^k(n) \downarrow 0$ as $n \uparrow \infty$, for each m, k, then there exists a subsequence, $a_m^{k_m}(n)$, such that*

$$\bigwedge_{n=1}^{\infty} \bigvee_{m=1}^{\infty} a_m^{k_m}(n) = 0.$$

Remark. A regular Boolean algebra is obviously complete.

THEOREM 6.1. *A necessary and sufficient condition that Ω be regular is that B be regular.*

Proof is omitted.

We state, also without giving the proof:

THEOREM 6.2. *A Boolean algebra of type B_μ is regular.*

We know therefore that the space of measurable functions on the unit interval is regular, while the subspace of bounded functions is not.

2. **Conjugate spaces.** Riesz [10] has developed the theory of additive functionals defined on a "fundamental domain." Such a fundamental domain is given by the positive and zero elements of any vector lattice. Riesz proves, in different language, that the space of additive functionals, each bounded above by a positive additive functional, defined on a vector lattice, is a σ-complete vector lattice. If the underlying space is σ-complete we can restrict ourselves to linear (continuous) functionals. In this case each functional is automatically bounded above by a positive linear functional, and we can state

THEOREM 6.3. *The space of linear functionals on a σ-complete vector lattice is a complete vector lattice.*

Proof is omitted.

The notion of conjugate space becomes particularly interesting to us, because of the possibility of multiplying elements of Ω, and because of the Nikodym theorem, in the case of the Λ^p spaces. Following closely the methods given by Banach [11] for the L^p spaces, we can prove the following theorems:

THEOREM 6.4. (Hölder's inequality.) *If $f\in\Lambda^p$, $g\in\Lambda^q$, $1/p+1/q=1$, then $fg\in\Lambda^1$. If f, $g\geqq 0$, then*

$$\int fg d\mu \leqq \left[\int f^p d\mu\right]^{1/p}\left[\int g^q d\mu\right]^{1/q}.$$

THEOREM 6.5. *Every linear functional $A(f)$ on Λ^p, $p>1$, is of the form*

$$A(f) = \int fg d\mu$$

where $g\in\Lambda^q$, $1/p+1/q=1$, and the norm of g in Λ^q is

$$\|g\| = \|A\|.$$

THEOREM 6.6. *Every linear functional $A(f)$ on Λ^1 is of the form*

$$A(f) = \int fg d\mu$$

where $g\in M$, the space of bounded functions, and

$$\|g\| = \|A\|.$$

These theorems apply immediately to any spaces of type F^p. Many other theorems regarding L^p spaces, such as those concerned with weak convergence (cf. Banach [11, p. 197]) can be carried over to the Λ^p spaces. Spaces of the type defined by Orlicz [4] can also be extended to spaces on a Boolean algebra of type B_μ.

Bibliography

1. Hans Freudenthal, *Teilweise geordnete Moduln*, Proceedings, Akademie van Wetenschappen te Amsterdam, vol. 39 (1936), pp. 641–651.
2. S. Saks, *Theory of the Integral*, 1937.
3. L. V. Kantorovitch, *Lineare halbgeordnete Räume*, Recueil Mathématique, Moscow, (n.s.), vol. 2 (1937), pp. 121–165.
4. W. Orlicz, *Über eine gewisse Klasse von Räumen vom Typus* B, Bulletin International de l'Académie Polonaise des Sciences et des Lettres, A, 1932, pp. 207–220.
5. C. Carathéodory, *Entwurf für eine Algebraisierung des Integralbegriffs*, Sitzungsberichte der Bayerischen Akademie der Wissenschaften, 1938, pp. 27–69.
6. M. H. Stone, *The theory of representations for Boolean algebras*, these Transactions, vol. 40 (1936), pp. 37–111.
7. Franz Wecken, *Abstrakte Integrale und fastperiodische Funktionen*, Mathematische Zeitschrift, vol. 45 (1939), pp. 377–404.
8. Garrett Birkhoff, *Lattice Theory*, American Mathematical Society Colloquium Publications, vol. 25, 1940.
9. H. F. Bohnenblust, *An axiomatic characterization of L_p-spaces*, Duke Mathematical Journal, vol. 6 (1940), pp. 627–640.
10. Frédéric Riesz, *Sur quelques notions fondamentales dans la théorie générale des opérations linéaires*, Annals of Mathematics, (2), vol. 41 (1940), pp. 174–206.
11. S. Banach, *Théorie des Opérations Linéaires*, 1932.

University of Minnesota,
Minneapolis, Minn.

ON THE NUMBER OF PARTITIONS OF A NUMBER INTO UNEQUAL PARTS[1]

BY

LOO-KENG HUA

1. **Introduction.** Let $q(n)$ be the number of partitions of an integer n into unequal parts, or into odd parts[2]. Then

$$f(x) = 1 + \sum_{n=1}^{\infty} q(n)x^n = (1 + x)(1 + x^2)(1 + x^3) \cdots \tag{1.1}$$
$$= \frac{1}{(1 - x)(1 - x^3)(1 - x^5) \cdots}.$$

Hardy and Ramanujan[3] indicated that by their fundamental analytic method one can obtain the following result:

$$q(n) = \frac{1}{2^{1/2}} \frac{d}{dn} J_0\left[i\pi\left\{\tfrac{1}{3}\left(n + \tfrac{1}{24}\right)\right\}^{1/2}\right]$$
$$+ 2^{1/2} \cos\left(\tfrac{2}{3}\pi n - \tfrac{1}{9}\pi\right) \frac{d}{dn} J_2\left[\tfrac{1}{3} i\pi\left\{\tfrac{1}{3}\left(n + \tfrac{1}{24}\right)\right\}^{1/2}\right] + \cdots$$
$$+ \text{to } [\alpha n^{1/2}] \text{ terms} + O(1)$$

where α is an arbitrary constant. This result is less satisfactory than that concerning the number $p(n)$ of partitions (unrestricted) of n, since in the latter case the error term approaches zero with increasing n. Recently Rademacher[4] obtained an equality for $p(n)$. The object of the present paper is to find an equality for $q(n)$. The work of this paper is a straightforward application of Hardy-Ramanujan's method with two modifications. These modifications are Kloosterman's sum and Rademacher's "Farey dissection of infinite order."

The present method may also be applied to find the explicit formula for

$$\sum_{x=1}^{[n^{1/2}]} p(n - x^2)$$

where $p(n)$ is the number of unrestricted partitions of n.

Presented to the Society, April 27, 1940; received by the editors January 9, 1941.

(1) This paper was accepted by Acta Arithmetica before the war.

(2) Cf. MacMahon, *Combinatory Analysis*, vol. 2, 1916, p. 11.

(3) Proceedings of the London Mathematical Society, (2), vol. 17 (1918), pp. 75–115.

(4) Proceedings of the London Mathematical Society, (2), vol. 43 (1937), pp. 241–254.

2. **Statement of the result.** Let

$$\epsilon_{h,k} = \begin{cases} \exp\left(-\pi i\left(\frac{(h'^2-1)}{8}\left(\frac{1-hh'}{k}-1\right)+\frac{h'(1-hh')}{8k} \right.\right. \\ \qquad \left.\left. +\frac{1}{24}\left(k+\frac{1-hh'}{k}\right)(hh'^2-h'-h)\right)\right), & \text{for } 2\mid k, \\ \exp\left(\frac{\pi i}{24}\left(k+\frac{1-hh'}{k}\right)(h+h'-h^2h')\right), & \text{for } 2\nmid k,\ 2\nmid h, \\ \exp\left(-\frac{\pi i}{8}\left(k^2-1-hk+\frac{1}{3}(h+h')\left(hh'k-\frac{hh'-1}{k}\right)\right)\right), & \\ & \text{for } 2\nmid k,\ 2\mid h, \end{cases}$$

and

$$\omega_{h,k} = \begin{cases} \epsilon_{h,k}\exp\left(-\frac{\pi i}{12k}(h+h')\right), & \text{for } 2\mid k, \\ \epsilon_{h,k}\exp\left(-\frac{\pi i}{24k}(2h-h')\right), & \text{for } 2\nmid k, \end{cases}$$

where $hh'\equiv 1 \pmod{k}$, $h\equiv h' \pmod{2}$.

Theorem. *The number of partitions of an integer n into unequal parts is given by*

$$q(n) = \frac{1}{2^{1/2}} \sum_{k=1,\,k \text{ odd}}^{\infty} \sum_{(h,k)=1,\,0<h\leq k} \omega_{h,k} e^{-2\pi i h n/k} \frac{d}{dn} J_0\left(\frac{i\pi}{k}\left\{\tfrac{2}{3}\left(n+\tfrac{1}{24}\right)\right\}^{1/2}\right),$$

where $J_0(x)$ is the Bessel function of the 0th order.

3. **Farey dissection.** By means of Cauchy's integral formula we obtain for (1.1)

$$q(n) = \frac{1}{2\pi i}\int_C \frac{f(x)}{x^{n+1}}\,dx.$$

The path of integration may be the circle defined as $|x| = e^{-2\pi N^{-2}}$ where N is a certain positive integer at our disposal. In the usual way we divide the circle into Farey arcs $\xi_{h,k}$ of order N. The Farey arc $\xi_{h,k}$ is defined by

$$x = \exp(2\pi i h/k - 2\pi N^{-2} + 2\pi i\vartheta), \qquad (h, k) = 1, \tag{3.1}$$

and

$$-\vartheta_1(h, k) = \frac{h+h_1}{k+k_1} - \frac{h}{k} \leq \vartheta \leq \frac{h+h_2}{k+k_2} - \frac{h}{k} = \vartheta_2(h, k) \tag{3.2}$$

where h_1/k_1, h/k, h_2/k_2 are three consecutive fractions in the Farey sequence of order N. It is well known that

$$(3.3)\qquad \frac{1}{k(N+k)} \leqq \vartheta_1(h, k) < \frac{1}{k(N+1)}, \quad \frac{1}{k(N+k)} \leqq \vartheta_2(h, k) < \frac{1}{k(N+1)}.$$

We obtain then

$$(3.4)\qquad q(n) = \frac{1}{2\pi i} \sum_{(h,k)=1,0<h\leqq k\leqq N} \int_{\xi_{h,k}} \frac{f(x)}{x^{n+1}}\, dx.$$

Let I_1 and I_2 denote the sums of those terms satisfying $2\mid k$, and $2\nmid k$, respectively. Then, by (3.4), we have

$$(3.5)\qquad q(n) = I_1 + I_2.$$

4. Lemmas on Kloosterman's sums.

LEMMA 4.1(5). *Let*

$$g(N, \vartheta, h, k) = \begin{cases} 1 & \text{for} \quad -\vartheta_1(h, k) \leqq \vartheta \leqq \vartheta_2(h, k), \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$g = \sum_{r=1}^{k} b_r e^{2\pi i r h'/k}$$

where h' is an integer satisfying

$$hh' \equiv 1 \pmod{k},$$

and b_r is independent of h and

$$\sum_{r=1}^{k} | b_r | < \log 4k.$$

LEMMA 4.2. *Let a be an absolute constant. Then*

$$\sum_{0<h\leqq ak,(h,ak)=1,h\equiv l(a)} \exp\left(\frac{2\pi i}{ak}(nk + mh')\right) = O(k^{2/3+\epsilon}(n, k)^{1/3}).$$

LEMMA 4.3. *If k is even and $\omega_{h,k}$ as defined in §2, then*

$$S_k = \sum_{1\leqq h\leqq k,(h,k)=1,hh'\equiv 1\,(k)} \omega_{h,k} e^{2\pi i(nh+mh')/k} = O(k^{2/3+\epsilon}(n, k)^{1/3}).$$

Proof. For the sake of simplicity I give here only the proof of the case $24\, k$.

(5) T. Estermann, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 7 (1929), pp. 93, 94.

Then

$$S_k = \sum_{1 \le l \le 24,\, (l,24)=1} \; \sum_{1 \le h \le k,\, (h,k)=1,\, hh' \equiv 1,\, h \equiv l \ (24)} \omega_{h,k} e^{2\pi i (nh + mh')/k}.$$

The inner sum becomes a Kloosterman's sum as in Lemma 4.2. Therefore we have

$$S_k = O(k^{2/3+\epsilon}(n, k)^{1/3}).$$

As to the proof of the other cases, nothing is difficult but a little complicated, and the following fact is used: let

$$F(h, k) = \omega_{h,k} e^{2\pi i (nh + mh')/k};$$

then $F(h+k, k) = F(h, k)$.

LEMMA 4.4. *Let* $2 \nmid k$ *and* $\omega_{h,k}$ *be as defined in* §2, *then*

$$S = \sum_{1 \le h \le k,\, (h,k)=1,\, hh' \equiv 1 \ (k),\, h' \text{ odd}} \omega_{h,k} e^{\pi i (2nh + mh')/k} = O(k^{2/3+\epsilon}(h, k)^{1/3}).$$

The proof is similar to that of Lemma 4.3, only notice that

$$S = \sum_{1 \le h < 2k,\, (h,2k)=1,\, hh' \equiv 1 \ (2k)} .$$

5. Lemmas from the theory of the linear transformation of the elliptic modular functions.

LEMMA 5.1. *Suppose that* $2 \nmid h$, $2 \mid k$; *that* h' *is a positive integer satisfying* $hh' \equiv 1 \pmod{k}$; *that* $\omega_{h,k}$ *is defined in* §2; *and that*

$$x = \exp\left(-\frac{2\pi z}{k} + \frac{2h\pi i}{k}\right), \qquad x' = \exp\left(-\frac{2\pi}{kz} - \frac{2h'\pi i}{k}\right),$$

where the real part of z *is positive. Then*

$$f(x) = \omega_{h,k} \exp\left(-\frac{\pi}{12kz} + \frac{\pi z}{12k}\right) f(x').$$

Proof. If we take $a = h$, $b = -k$, $c = (1 - hh')/k$, $d = h'$, so that $ad - bc = 1$, and write

$$x = q^2 = e^{2\pi i \tau}, \qquad x' = Q^2 = e^{2\pi i T},$$
$$\tau = (h + iz)/k, \qquad T = (-h' + i/z)/k,$$

then we can easily verify that

$$T = \frac{c + d\tau}{a + b\tau}.$$

Also, in the notation of Tannery and Molk, we obtain

$$f(x) = \frac{1}{2^{1/3}} q^{-1/12} \frac{\phi(\tau)}{\chi(\tau)}, \qquad f(x') = \frac{1}{2^{1/3}} Q^{-1/12} \frac{\phi(T)}{\chi(T)}.$$

Then

$$\begin{aligned} f(x') &= \frac{1}{2^{1/3}} Q^{-1/12} \frac{\phi(T)}{\chi(T)} = \exp\left(\pi i\left(\tfrac{1}{8}(d^2-1)(c-1) + \frac{cd}{8}\right.\right. \\ &\qquad \left.\left. - \frac{(b-c)(bcd-a)}{24}\right)\right) \frac{1}{2^{1/3}} Q^{-1/12} \frac{\phi(\tau)}{\chi(\tau)} \\ &= \exp\left(\pi i\left(\tfrac{1}{8}(d^2-1)(c-1) + \frac{cd}{8} - \frac{(b-c)(bcd-a)}{24}\right)\right) q^{1/12} Q^{-1/12} f(x) \\ &= \exp\left(\pi i\left(\tfrac{1}{8}(d^2-1)(c-1) + \frac{cd}{8} - \frac{(b-c)(bcd-a)}{24}\right)\right) \\ &\qquad \cdot \exp\left(\frac{\pi}{12k}\left(\frac{1}{z} - z\right)\right) \exp\left(\frac{\pi i}{12k}(h+h')\right) f(x). \end{aligned}$$

LEMMA 5.2. *Suppose that* $2 \nmid hk$ *and* $hh' \equiv 1 \pmod{2k}$, *that*

$$f_1(x) = \prod_1^\infty (1 + x^{n-1/2}) = 1 + \sum_{n=1}^\infty q_1(n) x^{n/2}.$$

Then

$$f(x) = \frac{\omega_{h,k}}{2^{1/2}} \exp\left(\frac{\pi}{12k}\left(z + \frac{1}{2z}\right)\right) f_1(x').$$

Proof. As in Lemma 5.1, we have

$$\begin{aligned} f_1(x) &= f_1(q^2) = \prod (1 + q^{2n-1}) = 2^{1/6} q^{1/24} \frac{1}{\chi(\tau)}, \\ f_1(x') &= 2^{1/6} Q^{1/24} \frac{1}{\chi(T)} = 2^{1/6} Q^{1/24} \exp\left(-\frac{(b-c)(abc-d)}{24}\pi i\right) \frac{\phi(\tau)}{\chi(\tau)} \\ &= 2^{1/6} Q^{1/24} \exp\left(-\frac{(b-c)(abc-d)}{24}\pi i\right) 2^{1/3} q^{1/12} f(x) \\ &= \exp\left(-\frac{(b-c)(abc-d)}{24}\pi i\right) 2^{1/2} \\ &\qquad \cdot \exp\left(\frac{\pi i}{24}\left(-\frac{h'}{k} + \frac{i}{kz} + \frac{2h}{k} + \frac{2iz}{k}\right)\right) f(x). \end{aligned}$$

LEMMA 5.3. *Suppose that* $2 \mid h$, $2 \nmid k$, $hh' \equiv 1 \pmod{k}$, $2 \mid h'$ *and suppose that*

$$f_2(x) = \prod_1^\infty (1 - x^{n-1/2}) = 1 + \sum q_2(n) x^{n/2}.$$

Then

$$f(x) = \frac{\omega_{h,k}}{2^{1/2}} \exp\left(\frac{\pi}{12k}\left(z + \frac{1}{2z}\right)\right) f_2(x').$$

Proof. We take

$$a = -h, \qquad b = k, \qquad c = (hh' - 1)/k, \qquad d = -h'.$$

Then

$$\begin{aligned} f_2(x') = f_2(Q^2) &= 2^{1/6} Q^{1/24} \frac{\psi(T)}{\chi(T)} \\ &= 2^{1/6} Q^{1/24} \exp\left(\frac{\pi i}{2}\left(\frac{b^2-1}{4} + \frac{ab}{4} - \frac{(a+d)(abd-c)}{12}\right)\right) \frac{\phi(\tau)}{\chi(\tau)} \\ &= 2^{1/2} \exp\left(\frac{\pi i}{2}\left(\frac{b^2-1}{4} + \frac{ab}{4} - \frac{(a+d)(abd-c)}{12}\right)\right) Q^{1/24} q^{1/12} f(x). \end{aligned}$$

6. **Approximation of the integrand.** Let

$$z = k(N^{-2} - i\vartheta).$$

Then

$$\begin{aligned} I_1 &= \sum_{1 \leqq k \leqq N, 2 \mid k} \; \sum_{(h,k)=1, 0<h<k} \int_{-k^{-1}(N+1)^{-1}}^{k^{-1}(N+1)^{-1}} g(\vartheta) f(e^{(2\pi i h - 2\pi z)/k}) e^{-2\pi i h n/k + 2\pi z n/k} d\vartheta \\ &= \sum_{1 \leqq k \leqq N, 2 \mid k} \; \sum_{(h,k)=1, 0<h<k} \int_{-k^{-1}(N+1)^{-1}}^{k^{-1}(N+1)^{-1}} g(\vartheta) \omega_{h,k} e^{(\pi/12k)(z-1/z)} \\ &\qquad \cdot f(x') e^{-2\pi i h n/k + 2\pi z n/k} d\vartheta \\ &= \sum_{1 \leqq k \leqq N, 2 \mid k} \; \sum_{(h,k)=1, 0<h<k} \int_{-k^{-1}(N+1)^{-1}}^{k^{-1}(N+1)^{-1}} g(\vartheta) \omega_{h,k} \\ &\qquad \cdot e^{(\pi/12k)(z-1/z) - 2\pi i h n/k + 2\pi z n/k} \sum_{\nu=0}^{\infty} q(\nu) e^{-(2\pi/kz + 2h'\pi i/k)\nu} d\vartheta \qquad (6.1) \\ &= \sum_{1 \leqq k \leqq N, 2 \mid k} \; \sum_{(h,k)=1, 0<h<k} \int_{-k^{-1}(N+1)^{-1}}^{k^{-1}(N+1)^{-1}} \sum_{\nu=0}^{\infty} q(\nu) e^{-(2\pi/kz)(\nu+1/24) + (2\pi z/k)(n+1/24)} \\ &\qquad \cdot \sum_{r=1}^{k} b_r e^{2\pi i r h'/k} \omega_{h,k} e^{-2\pi i h n/k - 2\pi i h' \nu/k} d\vartheta. \end{aligned}$$

Since $(1/k)\Re(1/z) \geqq \frac{1}{2}$, we have

$$
\begin{aligned}
|I_1| \leqq & \sum_{1\leqq k\leqq N, 2|k} \int_{-k^{-1}(N+1)^{-1}}^{k^{-1}(N+1)^{-1}} \sum_{\nu=0}^{\infty} q(\nu) \\
& \cdot \exp\left\{-\frac{2\pi}{k}\left(\nu+\frac{1}{24}\right)\Re\frac{1}{z}+\frac{2\pi}{k}\left(n+\frac{1}{24}\right)\Re z\right\} \\
& \cdot \sum_{r=1}^{k} |b_r| \left| \sum_{(h,k)=1} \omega_{h,k} e^{-2\pi i h n/k + 2h'(r-\nu)\pi i/k} \right| d\vartheta \\
& = O\left(\sum_{k=1}^{N} \int_{-k^{-1}(N+1)^{-1}}^{k^{-1}(N+1)^{-1}} \sum_{\nu=0}^{\infty} q(\nu) e^{-\pi(\nu+1/24)} \sum_{r=1}^{k} |b_r| k^{2/3} d\vartheta\right) \\
& = O\left(\sum_{k=1}^{N} \log k \cdot k^{2/3} \frac{1}{kN}\right) = O\left(\frac{1}{N}\sum_{k=1}^{N} k^{-1/3+\epsilon}\right) \\
& = O(N^{-1/3+\epsilon}).
\end{aligned}
$$

Let

$$
\begin{aligned}
J = \frac{1}{2^{1/2}} \sum_{k=1, k \text{ odd}}^{N} \sum_{(h,k)=1, 0<h\leqq k} & \int_{-k^{-1}(N+1)^{-1}}^{k^{-1}(N+1)^{-1}} g(\vartheta)\omega_{h,k} \\
& \cdot e^{(\pi/24k)(2z+1/z)-2\pi i h n/k + 2\pi z n/k} d\vartheta.
\end{aligned}
$$

The same method will give us that $|I_2 - J| = O(N^{-1/3+\epsilon})$.

7. **A contour integration.** Let $w = N^{-2} - i\vartheta$. Then

$$
\begin{aligned}
J &= \frac{-i}{2^{1/2}} \sum_{1\leqq k\leqq N, k \text{ odd}} \sum_{(h,k)=1, 0<h\leqq k} \omega_{h,k} e^{-2\pi i h n/k} \int_{N^{-2}-i\vartheta_2}^{N^{-2}+i\vartheta_1} e^{2\pi w(n+1/24)+\pi/24k^2 w} dw \\
&= \frac{i}{2^{1/2}} \sum_{1\leqq k\leqq N, k \text{ odd}} \sum_{(h,k)=1, 0<h\leqq k} \omega_{h,k} e^{-2\pi i h n/k} \left(\int_{N^{-2}+i\vartheta_1}^{N^{-2}+ik^{-1}(N+1)^{-1}} \right. \\
&\quad + \int_{N^{-2}+ik^{-1}(N+1)^{-1}}^{-N^{-2}+ik^{-1}(N+1)^{-1}} + \int_{-N^{-2}+ik^{-1}(N+1)^{-1}}^{-N^{-2}-ik^{-1}(N+1)^{-1}} + \int_{-N^{-2}-ik^{-1}(N+1)^{-1}}^{N^{-2}+ik^{-1}(N+1)^{-}} \\
&\quad \left. + \int_{N^{-2}-ik^{-1}(N+1)}^{N^{-2}-i\vartheta_2} - 2\pi i \text{ Residue at } 0 \right) \\
&= K_1 + K_2 + K_3 + K_4 + K_5 + L \text{ (say)}.
\end{aligned}
$$

We have

$$
\begin{aligned}
K_1 = \frac{i}{2^{1/2}} & \sum_{1\leqq k\leqq N, k \text{ odd}} \sum_{(h,k)=1, 0<h\leqq k} \omega_{h,k} e^{-2\pi i h n/k} \\
& \cdot \int_{N^{-2}+ik^{-1}(N+k)^{-1}}^{N^{-2}+ik^{-1}(N+1)^{-1}} g(\vartheta) e^{2\pi w(n+1/24)+\pi/24k^2 w} dw.
\end{aligned}
$$

By Lemma 3.1, we have

$$K_1 = O\left(\sum_{1 \le k \le N,\, k \text{ odd}} k^{2/3+\epsilon} \int_{k^{-1}(N+k)^{-1}}^{k^{-1}(N+1)^{-1}} \exp\left\{ 2\pi\left(n + \frac{1}{24}\right)\Re w + \frac{\pi}{24k^2} \Re \frac{1}{w} \right\} dw \right.$$
$$= O\left(\sum_{k=1}^{N} k^{2/3+\epsilon} e^{-2\pi n N^{-2}} \int_{k^{-1}(N+k)^{-1}}^{k^{-1}(N+1)^{-1}} d\vartheta \right)$$
$$= O(N^{-1/3+\epsilon}).$$

Similar result holds for K_5.

We have

$$\Re \frac{1}{k^2 w} = \frac{N^{-2}}{k^2 N^{-2} + N^2}, \qquad K_2 = O\left(\sum_{k=1}^{N} N^{-2} k^{2/3+\epsilon} \right) = O(N^{-1/3+\epsilon}).$$

Similar result holds for K_4.

Applying again Kloosterman's argument to K_3, we have also $K_3 = O(N^{-1/3})$.

Finally we find the residue of $\exp\,(2\pi w(n+1/24) + \pi/24k^2 w)$ at $w=0$. We have the expansion

$$e^{2\pi w(n+1/24)} = \sum_{\nu=1}^{\infty} \frac{(2\pi w(n + 1/24))^{\nu}}{\nu!},$$
$$e^{\pi/24k^2 w} = \sum_{\mu=1}^{\infty} \frac{1}{\mu!}\left(\frac{\pi}{24k^2 w}\right)^{\mu}.$$

The residue is, therefore,

$$\sum_{\mu=1}^{\infty} \frac{1}{\mu!}\left(\frac{\pi}{24k^2}\right)^{\mu} \frac{1}{(\mu-1)!} (2\pi(n + \tfrac{1}{24}))^{\mu-1}$$
$$= \frac{1}{2\pi}\frac{d}{dn}\sum_{\mu=1}^{\infty} \frac{1}{(\mu!)^2}\left(\frac{\pi}{24k^2}\right)^{\mu}(2\pi(n + \tfrac{1}{24}))^{\mu}$$
$$= \frac{1}{2\pi}\frac{d}{dn}\sum_{\mu=1}^{\infty} \frac{1}{2^{2\mu}(\mu!)^2}\left(\frac{\pi}{k}\,\{\tfrac{1}{3}(n + \tfrac{1}{24})\}^{1/2}\right)^{2\mu}$$
$$= \frac{1}{2\pi}\frac{d}{dn} J_0\left(\frac{i\pi}{k}\,\{\tfrac{1}{3}(n + \tfrac{1}{24})\}^{1/2}\right).$$

Therefore

$$q(n) = \frac{1}{2^{1/2}} \sum_{k=1,\, k \text{ odd}}^{N} \sum_{(h,k)=1,\, 0<h\le k} \omega_{h,k} e^{-2\pi i h n/k} \frac{d}{dn} J_0\left(\frac{i\pi}{k}\,\{\tfrac{1}{3}(n + \tfrac{1}{24})\}^{1/2}\right)$$
$$+ O(N^{-1/3+\epsilon}).$$

Let $N \to \infty$; we obtain the theorem.

NATIONAL TSING HUA UNIVERSITY,
KUNMING, YÜNNAN, CHINA

THEORY OF REDUCTION FOR ARITHMETICAL EQUIVALENCE. II[1]

BY

HERMANN WEYL

1. Introduction. Lattices over the unit lattice. Given n linearly independent vectors $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$ in an n-dimensional vector space E^n, the formula

$$(1) \qquad \mathfrak{x} = y_1\mathfrak{b}_1 + \cdots + y_n\mathfrak{b}_n$$

yields all vectors of the space E^n or of a lattice $\mathfrak{L}$ in E^n if the coordinates y_i range over all real numbers or all integers, respectively. We take the viewpoint that the lattice $\mathfrak{L}$ is given but the choice of its basis arbitrary. The several bases are connected with one another by unimodular transformations. If $f(\mathfrak{x})$ is a gauge function assigning a "length" $f(\mathfrak{x})$ to each vector $\mathfrak{x}$ the problem of reduction requires normalization of the lattice basis in terms of the given f. A solution is sought for all possible gauge functions or at least for some important class. The most significant class is obtained by running f^2 over all positive quadratic forms.

Following in Dirichlet's and Hermite's footsteps, Minkowski developed such a method of reduction for quadratic forms and established the decisive facts about it. In R1 I approached the same problem in that geometric way which Minkowski had initiated but then abandoned for unknown reasons.

The question may be put in a slightly different form. All linear mappings of E^n carrying $\mathfrak{L}$ into itself carry $f(\mathfrak{x})$ into equivalent gauge functions. The task is to pick out by a universal rule in each class of equivalent gauge functions one particular $f(\mathfrak{x})$ which is called the reduced function of its class. Let $\mathfrak{R}_0$, $\mathfrak{R}$, $\mathfrak{C}$ in the future denote the fields of all rational, real and complex numbers, respectively. Complex numbers are written in the form $\xi = x_0 + x_1 i$ (x_0, x_1 real). It is convenient to insert between the full vector space and the lattice $\mathfrak{L}$, the set E_0^n of all vectors (1) with *rational* coefficients y_i, a set which we describe as an n-dimensional vector space over $\mathfrak{R}_0$. Crystallography has found this advisable in distinguishing between the macroscopic and atomistic symmetries of a crystal, and in the theory of algebraic numbers one puts the *field* before the *ring* of its integers.

Let a lattice $\mathfrak{L}$ in E_0^n be given. With respect to any basis $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$ of E_0^n, formula (1), the function $f(\mathfrak{x})$ is represented by a function $g(y_1, \cdots, y_n)$ and the lattice $\mathfrak{L}$ by a "numerical lattice" Λ whose vectors are n-uples $(y_1, \cdots, y_n)$

Presented to the Society, January 1, 1941; received by the editors December 11, 1940.

[1] The first part, which appeared under the same title in these Transactions, vol. 48 (1940), pp. 126–164, is cited as R1.

of rational numbers. (Only if $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$ is a true basis of $\mathfrak{L}$ will Λ be the unit lattice I whose elements are the n-uples of integers.) Hence $f(\mathfrak{x})$ with respect to $\mathfrak{L}$ is represented by g/Λ. All representations g/Λ of $f(\mathfrak{x})/\mathfrak{L}$ are equivalent, i.e., they arise from one another by linear transformations of the coordinates with rational coefficients. In each class of equivalent g/Λ we are to pick one individual, the "reduced" g/Λ. Suppose we have succeeded in doing this by some universal rule. We then have to select, for each Λ that may occur in a reduced g/Λ, a definite basis $\mathfrak{b}_1^*, \cdots, \mathfrak{b}_n^*$ in terms of which $\mathfrak{L}$ is represented by Λ. The equation

$$f^*(\mathfrak{x}) = g(y_1, \cdots, y_n) \quad \text{for} \quad \mathfrak{x} = y_1\mathfrak{b}_1^* + \cdots + y_n\mathfrak{b}_n^*$$

then defines the reduced gauge function f^* in its class. By the first step of reducing g/Λ no essential progress has been made unless the lattices Λ which may occur in a reduced g/Λ are limited to a finite number of possibilities. For only then is the selection of a basis $\mathfrak{b}_1^*, \cdots, \mathfrak{b}_n^*$ for each of these Λ essentially simpler than the original problem.

The Dirichlet-Hermite-Minkowski method of reduction by admitting only bases $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$ of $\mathfrak{L}$ always represents $\mathfrak{L}$ by the one lattice $\Lambda = \mathrm{I}$, the unit lattice. Thus it provides the ideal solution. Minkowski's construction of consecutive shortest distances in the lattice

$$f(\mathfrak{b}_1) = M_1, \cdots, f(\mathfrak{b}_n) = M_n$$

(for which he obtains the inequality $M_1 \cdots M_n V \leq 2^n$) falls under our more general scheme. That theorem which he describes as indicating a certain "Oekonomie der Strahldistancen" states exactly that there is only an *a priori* limited number of possibilities for Λ with which to count in a reduced g/Λ. In R1 I carried the first method over to those other fields and quasi-fields which have not more than one infinite prime spot, and I found that it works only under the hypothesis that the class number for ideals is 1. Simultaneously Siegel observed that the rougher second method, by which incidentally Minkowski had proved that the class number of positive quadratic forms with *integral* coefficients and a given discriminant is finite, operates without this restrictive hypothesis(²). I add the remark that an argument making no use of the bases of a lattice need not even assume their existence. In an algebraic number field $\mathcal{J}$ we consider any "order" $[\mathcal{J}]$; in general there are several classes of lattices belonging to this order. The theory is limited neither to the principal class nor to the principal order. Following a suggestion by Siegel, P. Humbert generalized the investigation of quadratic forms to an arbitrary algebraic number field $\mathcal{J}$ with several infinite prime spots(³). No doubt the whole problem thereby loses much of its simplicity. But once upon this track one ought to include the quaternions and thus deal also with those noncom-

(³) See P. Humbert, Commentarii Mathematici Helvetici, vol. 12 (1939–1940), pp. 263–306.

mutative division algebras of finite degree over $\mathfrak{R}_0$ for which the concept of infinite prime spots goes through. I resume here the rougher method of reduction with these further generalizations by the same geometric approach as in R1. I am not only interested in the fact that certain numbers are finite; I wish to ascertain reasonably low explicit upper bounds for them. The geometric method yields good results in this regard.

Before concluding this introduction I remind the reader of some simple facts about lattices in E_0^n. A vector $\mathfrak{x}$ in E_0^n is defined as an n-uple $(x_1, \cdots, x_n)$ of rational numbers. The unit vectors $\mathfrak{e}_k = (e_{1k}, \cdots, e_{nk})$ are the columns of the unit matrix $\|e_{ik}\|$. The word lattice means any set of vectors such that $\mathfrak{a} - \mathfrak{b}$ is contained in the set every time $\mathfrak{a}$ and $\mathfrak{b}$ are. We assume that the lattice is n-dimensional, i.e., contains n linearly independent vectors; and discrete, i.e., we require that for any given positive integer q there are not more than a finite number of lattice vectors satisfying the inequalities

$$|x_1| \leqq q, \cdots, |x_n| \leqq q.$$

From now on the term lattice refers only to discrete lattices which have the full dimensionality of their vector space. By a familiar argument one shows that one can find n linearly independent vectors $\mathfrak{l}_1, \cdots, \mathfrak{l}_n$ in a given lattice $\mathfrak{L}$ such that every lattice vector

$$\mathfrak{x} = u_1\mathfrak{l}_1 + \cdots + u_n\mathfrak{l}_n$$

has integral components u_i. By the same construction one adapts the basis $\mathfrak{l}_1, \cdots, \mathfrak{l}_n$ of any lattice Λ containing the unit lattice I to the basis $\mathfrak{e}_1, \cdots, \mathfrak{e}_n$ of I:

$$\begin{aligned} \mathfrak{e}_1 &= c_1\mathfrak{l}_1, \\ \mathfrak{e}_2 &= c_{21}\mathfrak{l}_1 + c_2\mathfrak{l}_2, \\ &\cdots\cdots\cdots\cdots\cdots, \\ \mathfrak{e}_n &= c_{n1}\mathfrak{l}_1 + \cdots + c_{n,n-1}\mathfrak{l}_{n-1} + c_n\mathfrak{l}_n. \end{aligned}$$

The integers c_k are positive and the integral skew coefficients c_{ki} $(i<k)$ may be normalized by

$$0 \leqq c_{ki} < c_i \qquad\qquad (k = i + 1, \cdots, n);$$

then $(\mathfrak{l}_1, \cdots, \mathfrak{l}_n)$ is uniquely determined. The index $j = [\Lambda : \mathrm{I}]$, i.e., the number of vectors in Λ which are incongruent modulo I, equals $c_1 \cdots c_n$. Let $\Lambda^{(k)}$ denote the part of Λ lying in the linear subspace $x_{k+1} = \cdots = x_n = 0$. The index $j_k = [\Lambda^{(k)} : \mathrm{I}^{(k)}]$ equals $c_1 \cdots c_k$. Hence these two lemmas:

LEMMA 1. *j_k is a divisor of j_h for $k<h$.*

LEMMA 2. *The number $h_n(j)$ of different lattices Λ over* I *of given index $j = [\Lambda : \mathrm{I}]$ is finite.*

Indeed, it equals the sum

$$\sum c_1^{n-1} c_2^{n-2} \cdots c_n^{0}$$

extended over all factorizations $c_1c_2 \cdots c_n = j$ of j. (Incidentally, the numbers $h_n(j)$ for $j=1, 2, \cdots$ have as their generating function the Dirichlet series

$$\sum_{j=1}^{\infty} h_n(j)\cdot j^{-s} = \zeta(s)\zeta(s-1)\cdots\zeta(s-n+1)$$

convergent in the half-plane $\Re s > n$.)

2. **Vector space and lattice over an algebraic field.** Let $\mathfrak{J}$ be any field of finite degree f over $\Re_0$. By carefully putting all factors in their proper places we shall see to it that all arguments and formulas in this and the following two sections remain valid for any division algebra, whether commutative or not, of finite degree over $\Re_0$. We choose a basis $\sigma_1, \cdots, \sigma_f$ of $\mathfrak{J}/\Re_0$ so that each number ξ of $\mathfrak{J}$ is uniquely represented by

$$\xi = x_1\sigma_1 + \cdots + x_f\sigma_f \qquad (x_\alpha \text{ rational}). \tag{2}$$

Any n-uple $(\xi_1, \cdots, \xi_n)$ of numbers ξ_i in $\mathfrak{J}$ is a vector of the n-dimensional vector space E^n over $\mathfrak{J}$. The fundamental operations are addition of two vectors, $\mathfrak{x}+\mathfrak{x}'$, and multiplication $\delta\mathfrak{x}$ of a vector $\mathfrak{x}$ by a number δ (the numerical factor always in front of the vector!). Thus we may write

$$\mathfrak{x} = (\xi_1, \cdots, \xi_n) = \xi_1\mathfrak{e}_1 + \cdots + \xi_n\mathfrak{e}_n.$$

k linearly independent vectors $\mathfrak{d}_1, \cdots, \mathfrak{d}_k$ span a linear subspace $[\mathfrak{d}_1, \cdots, \mathfrak{d}_k]$ consisting of all vectors of the form $\eta_1\mathfrak{d}_1 + \cdots + \eta_k\mathfrak{d}_k$. Any n linearly independent vectors $\mathfrak{d}_1, \cdots, \mathfrak{d}_n$ form a basis of $E^n/\mathfrak{J}$ in terms of which each vector is uniquely expressible as

$$\mathfrak{x} = \eta_1\mathfrak{d}_1 + \cdots + \eta_n\mathfrak{d}_n. \tag{3}$$

The original coordinates ξ_i are connected with the η_i by that nonsingular linear transformation D,

$$\xi_i = \sum_k \eta_k\delta_{ik}, \tag{4}$$

whose matrix $\|\delta_{ik}\|$ has for its columns the vectors $\mathfrak{d}_k = (\delta_{1k}, \cdots, \delta_{nk})$.

Expressing each component ξ_i in terms of the basis σ of $\mathfrak{J}$,

$$\xi_i = x_{i1}\sigma_1 + \cdots + x_{if}\sigma_f,$$

we identify $E^n/\mathfrak{J}$ with the (nf)-dimensional vector space E_0^{nf} over $\Re_0$. The rational numbers $x_{i\alpha}$ are the coordinates of $\mathfrak{x}$ with respect to the basis $\sigma_\alpha\mathfrak{e}_i$. One has to distinguish between linear dependence in $\mathfrak{J}$ and in $\Re_0$.

We suppose we are given a lattice $\mathfrak{L}$ in E_0^{nf}. It will have a basis $\mathfrak{l}_\alpha$

($\mu=1, \cdots, nf$) in terms of which each vector $\mathfrak{x}$ of $\mathfrak{L}$,

$$\mathfrak{x} = \sum_{\mu} u_\mu \mathfrak{l}_\mu \tag{5}$$

has rational *integral* components u_μ. A number δ of $\mathcal{J}$ is said to be a *multiplier* of $\mathfrak{L}$ if the operation $\mathfrak{x} \to \delta\mathfrak{x}$ carries each lattice vector $\mathfrak{x}$ into a lattice vector $\delta\mathfrak{x}$. The multipliers of $\mathfrak{L}$ form an *order* $[\mathcal{J}]$. This assertion is meant to imply the following four properties(3):

1°. The number 1 is in $[\mathcal{J}]$.

2°. $[\mathcal{J}]$ is a ring.

3°. Any given number δ in $\mathcal{J}$ may be multiplied by a positive rational integer m such that $m\delta$ is in $[\mathcal{J}]$.

4°. Each number in $[\mathcal{J}]$ is an integer.

1° and 2° are evident. To prove 3° and 4° we write

$$\delta\mathfrak{l}_\mu = \sum_{\nu} d_{\nu\mu}\mathfrak{l}_\nu \qquad (d_{\mu\nu} \text{ rational}).$$

If δ is any number and m a common denominator of the coefficients $d_{\mu\nu}$, then $m\delta$ is a multiplier. If δ happens to be a multiplier, then the $d_{\mu\nu}$ are rational integers and δ satisfies the equation

$$| \delta e_{\nu\mu} - d_{\nu\mu} | = 0.$$

In the same manner as for the "principal order" consisting of all integers of $\mathcal{J}$ one proves(4) that any order $[\mathcal{J}]$ is a discrete f-dimensional lattice in the f-dimensional vector space $\mathcal{J}/\mathfrak{R}_0$, and hence has a basis $\sigma_1, \cdots, \sigma_f$ in terms of which every number ξ of $[\mathcal{J}]$ appears in the form (2) with rational *integral* coefficients x_i.

The transformation D, (4), maps $\mathfrak{L}$ upon a lattice Λ: If $\mathfrak{x}=(\xi_1, \cdots, \xi_n)$ is in $\mathfrak{L}$, then $(\eta_1, \cdots, \eta_n)$ is in Λ, and *vice versa*. We call two lattices *equivalent* and admit them to the same *class* if one is carried into the other by a non-singular transformation D. The lattices Λ of one class express a given lattice $\mathfrak{L}$ in terms of different bases $(\mathfrak{d}_1, \cdots, \mathfrak{d}_n)$ of $E^n/\mathcal{J}$. Obviously two equivalent lattices have the same multipliers.

A lattice $\mathfrak{L}$ is said to belong to the order $[\mathcal{J}]$ if every number of that order is a multiplier of $\mathfrak{L}$. (For $n=1$ this notion coincides with that of an ideal in $[\mathcal{J}]$, and our classes of lattices with the classes of ideals.) Given an order $[\mathcal{J}]$, the n-uples $(\xi_1, \cdots, \xi_n)$ of numbers ξ_i in $[\mathcal{J}]$ form a lattice I which belongs to the order $[\mathcal{J}]$; we call it the *unit lattice* for $[\mathcal{J}]$. The lattices belonging to a given order $[\mathcal{J}]$ are distributed over a number of classes, the class of I being the principal class.

(3) Notion and name are due to Dedekind. Hilbert in his Zahlbericht introduced the word "ring" for this purpose, but since ring has now acquired a wider meaning I revert, in agreement with such authorities as Artin and Chevalley, to Dedekind's terminology.

(4) Cf. H. Weyl, *Algebraic Theory of Numbers*, Princeton, 1940, pp. 145–146.

Let $\sigma_1, \cdots, \sigma_f$ be a basis of $[\mathfrak{J}]$ and $\mathfrak{l}_\mu$ $(\mu=1, \cdots, nf)$ a basis of Λ. If Λ contains I the vectors $\sigma_a \mathfrak{e}_k$, which span I, are linear combinations of the $\mathfrak{l}_\mu$ with integral rational coefficients, and their absolute determinant, i.e., the absolute determinant of the transformation connecting the coordinates u_μ with the x_{ka} $(k=1, \cdots, n; a=1, \cdots, f)$ is the index $j=[\Lambda:\mathrm{I}]$.

Those vectors $(\xi_1, \cdots, \xi_n)$ in Λ for which $\xi_{k+1}=\cdots=\xi_n=0$ form a lattice $\Lambda^{(k)}$ in the k-dimensional space $E^k/\mathfrak{J}$ with the coordinates $\xi_1, \cdots, \xi_k$. Considering Λ as a lattice in E_0^{nf} and using the arrangement

$$x_{11}, \cdots, x_{1f}; x_{21}, \cdots, x_{2f}; \cdots$$

of the coordinates in E_0^{nf} one can apply Lemma 1 to kf and $(k+1)f$ instead of k and h and thus one derives a corresponding proposition in $\mathfrak{J}$ instead of $\mathfrak{R}_0$:

LEMMA 3. *In the row of indices*

$$j_k = [\Lambda^{(k)}:\mathrm{I}^{(k)}] \qquad (k = 1, \cdots, n) \tag{6}$$

each number is a divisor of its successor.

The set of vectors $(\xi_1, \cdots, \xi_n)$ in Λ outside $[\mathfrak{e}_1, \cdots, \mathfrak{e}_{k-1}]$, i.e., for which $(\xi_k, \cdots, \xi_n) \neq (0, \cdots, 0)$, will be denoted Λ_k. Thus Λ_k and $\Lambda^{(k-1)}$ are complements in Λ.

We have seen that the number of different lattices Λ over a given lattice I with a given index $[\Lambda:\Gamma]=j$ is finite, namely $h_{nf}(j)$. More exactly, one finds by the same argument that the number $H_f(j_1, \cdots, j_n)$ of different lattices Λ over I with given indices (6) has as its generating function the Dirichlet series of n variables $s_1, \cdots, s_n$:

$$Z_f(s_1 + s_2 + \cdots + s_n - nf)\cdot Z_f(s_2 + \cdots + s_n - (n-1)f) \cdots Z_f(s_n - f) = \sum_{j_1, \cdots, j_n} H_f(j_1, \cdots, j_n)\cdot j_1^{-s_1} \cdots j_n^{-s_n}$$

where

$$Z_f(s) = \zeta(s+1) \cdots \zeta(s+f).$$

Hence *a fortiori*:

LEMMA 4. *We have found upper bounds for the number of lattices Λ belonging to a given order $[\mathfrak{J}]$ which contain the unit vectors $\mathfrak{e}_1, \cdots, \mathfrak{e}_n$ and hence the unit lattice* I *for $[\mathfrak{J}]$ and which, moreover, have either a given index $j=[\Lambda:\mathrm{I}]$ or a given row of indices $j_1, \cdots, j_n$.*

3. **Preliminaries about reduction.** Suppose an order $[\mathfrak{J}]$ and a basis $\sigma_1, \cdots, \sigma_f$ of $[\mathfrak{J}]$ to be given. We consider a real-valued function

$$f(\mathfrak{x}) = f(\xi_1, \cdots, \xi_n)$$

which depends on a variable vector $\mathfrak{x}$ in $E^n/\mathfrak{J}$ and is positive except for $\mathfrak{x}=0$, and we assume:

(i_0) For each positive q one can ascertain a positive q' such that the inequality $f(\mathfrak{x})<q$ implies the nf inequalities

$$|x_{i\alpha}| < q' \quad (i = 1, \cdots, n; \alpha = 1, \cdots, f) \tag{7}$$

for the components $x_{i\alpha}$ of the ξ_i.

Let $\mathfrak{L}$ be a lattice belonging to the fixed order $[\mathcal{J}]$. n vectors $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$ of $\mathfrak{L}$ which are linearly independent with respect to $\mathcal{J}$ constitute a semi-basis of $\mathfrak{L}$. Because of the discreteness of $\mathfrak{L}$ there is but a finite number of vectors $\mathfrak{x}$ in $\mathfrak{L}$ satisfying the inequalities (7). Hence Minkowski's construction of consecutive minima of f in $\mathfrak{L}$ is applicable. It yields a semi-basis $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$ of $\mathfrak{L}$ such that

$$f(\mathfrak{x}) \geqq f(\mathfrak{b}_k) = M_k$$

for every vector $\mathfrak{x}$ in $\mathfrak{L}$ outside $[\mathfrak{b}_1, \cdots, \mathfrak{b}_{k-1}]$ (*reduced semi-basis*). Obviously

$$M_1 \leqq M_2 \leqq \cdots \leqq M_n. \tag{8}$$

The mapping

$$\mathfrak{x} = \eta_1\mathfrak{b}_1 + \cdots + \eta_n\mathfrak{b}_n \to (\eta_1, \cdots, \eta_n) \tag{3}$$

carries $f(\mathfrak{x})$ into a function $g(\eta_1, \cdots, \eta_n)$ and $\mathfrak{L}$ into a lattice Λ which contains the unit lattice I for $[\mathcal{J}]$. The function $g(\xi_1, \cdots, \xi_n)$ is *reduced with respect to* Λ, i.e.,

$$g(\xi_1, \cdots, \xi_n) \geqq g(e_{1k}, \cdots, e_{nk}) \tag{9}$$

whenever $(\xi_1, \cdots, \xi_n)$ is in Λ_k.

The M_k are uniquely determined by f and $\mathfrak{L}$; the situation is somewhat less favorable for $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$. Suppose $\mathfrak{b}_1', \cdots, \mathfrak{b}_n'$ is another set constructed according to our prescription. If M_k is actually lower than M_{k+1} then $[\mathfrak{b}_1', \cdots, \mathfrak{b}_k'] = [\mathfrak{b}_1, \cdots, \mathfrak{b}_k]$. (Analogues of Theorems 8 and 9 in R1.)

Being given n real numbers p_k,

$$1 \leqq p_1 \leqq \cdots \leqq p_n,$$

we say that the semi-basis $\mathfrak{b}_1', \cdots, \mathfrak{b}_n'$ of $\mathfrak{L}$ has the property $B(p_1, \cdots, p_n)$ if

$$f(\mathfrak{x}) \geqq \frac{1}{p_k} f(\mathfrak{b}_k')$$

for any vector $\mathfrak{x}$ in $\mathfrak{L}$ outside $[\mathfrak{b}_1', \cdots, \mathfrak{b}_{k-1}']$. Accordingly we ascribe the property $B(p_1, \cdots, p_n)$ to a function $g(\xi_1, \cdots, \xi_n)$ in conjunction with a lattice Λ over I if

$$g(\xi_1, \cdots, \xi_n) \geqq \frac{1}{p_k} \cdot g(e_{1k}, \cdots, e_{nk})$$

whenever $(\xi_1, \cdots, \xi_n)$ is in Λ_k.

If $\mathfrak{b}_k'$ is a semi-basis of $\mathfrak{L}$ with the property $B(p_1, \cdots, p_n)$, then

$$M_k' = f(\mathfrak{b}_k') \leqq p_k M_k.$$

(Analogue of Theorem 8_p.) Indeed, let $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$ be a reduced semi-basis of $\mathfrak{L}$, $f(\mathfrak{b}_k) = M_k$. At least one of the k linearly independent vectors $\mathfrak{b}_1, \cdots, \mathfrak{b}_k$, say $\mathfrak{b}_i$, lies outside $[\mathfrak{b}_1', \cdots, \mathfrak{b}_{k-1}']$; hence

$$f(\mathfrak{b}_i) \geqq \frac{1}{p_k} \cdot f(\mathfrak{b}_k')$$

or

$$M_k' \leqq p_k M_i \leqq p_k M_k.$$

With the same notations I maintain that $[\mathfrak{b}_1', \cdots, \mathfrak{b}_k'] = [\mathfrak{b}_1, \cdots, \mathfrak{b}_k]$ provided $M_{k+1} > p_k M_k$. (Analogue of Theorem 9_p.) Proof: Suppose that one of the vectors $\mathfrak{b}_1', \cdots, \mathfrak{b}_k'$, say $\mathfrak{b}_i'$, is not in $[\mathfrak{b}_1, \cdots, \mathfrak{b}_k]$. Then

$$f(\mathfrak{b}_i') \geqq M_{k+1}.$$

Vice versa, if all k numbers $M_1', \cdots, M_k'$ are less than M_{k+1} then $\mathfrak{b}_1', \cdots, \mathfrak{b}_k'$ lie in $[\mathfrak{b}_1, \cdots, \mathfrak{b}_k]$. The observation that $M_i' \leqq p_i M_i \leqq p_k M_k$ finishes the proof.

The notation of *properly reduced* bases depends on a given multiplicative group U of numbers ϵ in $\mathcal{J}$ and deals with functions f which satisfy the further condition:

(ii$_0$) $f(\epsilon\mathfrak{x}) = f(\mathfrak{x})$ (ϵ in U).

The semi-basis $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$ of $\mathfrak{L}$ is said to be properly reduced provided the inequality

$$f(\mathfrak{x}) > f(\mathfrak{b}_k)$$

holds with the sign $>$ for any vector $\mathfrak{x}$ of $\mathfrak{L}$ outside $[\mathfrak{b}_1, \cdots, \mathfrak{b}_{k-1}]$ except for the vectors of the special form

$$\mathfrak{x} = \epsilon\mathfrak{b}_k \qquad (\epsilon \text{ in } U).$$

Accordingly $g(\xi_1, \cdots, \xi_n)$ is properly reduced with respect to the lattice Λ over I if

$$g(\xi_1, \cdots, \xi_n) > g(e_{1k}, \cdots, e_{nk})$$

for all vectors $(\xi_1, \cdots, \xi_n)$ in Λ_k except the special vectors

$$\epsilon(e_{1k}, \cdots, e_{nk}) \qquad (\epsilon \text{ in } U).$$

With $\mathfrak{b}_k$ the vectors

$$\mathfrak{b}_k' = \epsilon_k \mathfrak{b}_k \qquad (\epsilon_k \text{ in } U)$$

form a reduced semi-basis of $\mathfrak{L}$ under the sole assumption that they lie in $\mathfrak{L}$.

Because $\mathfrak{x}=\epsilon_k\mathfrak{d}_k$ satisfies

$$f(\mathfrak{x}) = f(\mathfrak{d}_k), \quad \text{a fortiori} \quad f(\mathfrak{x}) \leqq f(\mathfrak{d}_k),$$

there is then, according to (i_0), only a finite number of possibilities for ϵ_k. We set

$$\eta_1\mathfrak{d}_1 + \cdots + \eta_n\mathfrak{d}_n = \mathfrak{x} = \eta_1'\mathfrak{d}_1' + \cdots + \eta_n'\mathfrak{d}_n'$$

and denote by Λ, Λ' the corresponding images of $\mathfrak{L}$:

$$(\eta_1, \cdots, \eta_n) \text{ in } \Lambda \cdot \rightleftarrows \cdot \mathfrak{x} \text{ in } \mathfrak{L} \cdot \rightleftarrows \cdot (\eta_1', \cdots, \eta_n') \text{ in } \Lambda'.$$

The "special transformation"

$$\eta_k = \eta_k'\epsilon_k \qquad (\epsilon_k \text{ in } U) \tag{10}$$

carries Λ into Λ'. We count in the same *family* any two lattices Λ and Λ' arising from each other by such a special transformation. Given the lattice Λ over I there is only a finite number of special transformations such that the transformed lattice Λ' also contains I. In particular, the group $\{J_\Lambda\}$ of all special transformations J_Λ leaving Λ invariant is finite. If h is its degree, one has $\epsilon_k^h=1$ $(k=1, \cdots, n)$ for each J_Λ; hence the ϵ_k are roots of unity in $\mathcal{J}$. The roots of unity in a field $\mathcal{J}$ form a finite cyclic group; in particular, if $\mathcal{J}$ has at least one real spot, the only such roots are ± 1. (However, in noncommutative division algebras the group of the roots of unity is, generally speaking, neither Abelian nor finite.)

The simple argument in R1, p. 136, shows:

If $\mathfrak{d}_1, \cdots, \mathfrak{d}_n$ *is a properly reduced semi-basis and* $\mathfrak{d}_1', \cdots, \mathfrak{d}_n'$ *any semi-basis of* $\mathfrak{L}$, *then the sequence of the values* $f(\mathfrak{d}_1), \cdots, f(\mathfrak{d}_n)$ *is lower than* $f(\mathfrak{d}_1'), \cdots, f(\mathfrak{d}_n')$. *If* $\mathfrak{d}_1', \cdots, \mathfrak{d}_n'$ *is reduced and* $\mathfrak{d}_1, \cdots, \mathfrak{d}_n$ *properly reduced, then*

$$\mathfrak{d}_1' = \epsilon_1\mathfrak{d}_1, \cdots, \mathfrak{d}_n' = \epsilon_n\mathfrak{d}_n \qquad (\epsilon_i \textit{ in } U).$$

4. Extension to the ground field $\mathfrak{R}$. Minkowski's inequality. So far the function $f(\mathfrak{x})$ has been defined merely for the vectors in the space $E^n/\mathcal{J}$. In order to introduce geometry we assign to the variables x_α in (2) arbitrary *real* values:

$$\xi^* = x_1\sigma_1 + \cdots + x_f\sigma_f. \tag{2*}$$

Sticking to the multiplication table of the basic elements σ_α, we thus extend $\mathcal{J}/\mathfrak{R}_0$ to a commutative algebra $\mathcal{J}^*$ over $\mathfrak{R}$. But only in the two cases treated in R1, where $\mathcal{J}$ is $\mathfrak{R}_0$ itself or an imaginary quadratic field over $\mathfrak{R}_0$, is $\mathcal{J}^*$ again a field. In general it is not. However, any n-uple $\mathfrak{x}^*=(\xi_1^*, \cdots, \xi_n^*)$ of elements ξ_i^* in $\mathcal{J}^*$ may be considered as a vector in an (nf)-dimensional vector space E^{nf} over $\mathfrak{R}$ with the real coordinates $x_{i\alpha}$:

$$\xi_i^* = x_{i1}\sigma_1 + \cdots + x_{if}\sigma_f.$$

We now assume $f(\mathfrak{x}^*)$ to be a *gauge function*, i.e., a continuous real-valued function in this space, having the following properties:

(i) $f(\mathfrak{x}^*) > 0$ except for $\mathfrak{x}^* = 0$.
(ii) $f(t\mathfrak{x}^*) = |t| \cdot f(\mathfrak{x}^*)$ for any real factor t.
(iii) $f(\mathfrak{x}_1^* + \mathfrak{x}_2^*) \leqq f(\mathfrak{x}_1^*) + f(\mathfrak{x}_2^*)$.

The gauge body

$$K:\ f(\xi_1^*, \cdots, \xi_n^*) < 1$$

and also the solid qK defined by $f(\xi_1^*, \cdots, \xi_n^*) < q$ are bounded; hence postulate (i_0) of the previous section is fulfilled. Let V^* be the volume of K computed in terms of the coordinates $x_{i\alpha}$.

Again we fix an order $[\mathfrak{J}]$ and a basis $\sigma_1, \cdots, \sigma_f$ of $[\mathfrak{J}]$. Let Λ be a lattice belonging to this order and containing the unit lattice I for $[\mathfrak{J}]$ and let $f(\xi_1, \cdots, \xi_n)$ be reduced with respect to Λ. The volume of K in terms of the coordinates u_μ as introduced by (5), i.e., measured against the fundamental parallelepiped of Λ, equals $V^* \cdot [\Lambda : \mathrm{I}]$. Hence by the simple argument explained in R1, p. 140, Minkowski's second inequality leads to this formula holding for a gauge function $f(\xi_1^*, \cdots, \xi_n^*)$ which is reduced with respect to Λ:

$$(M_1 \cdots M_n)^f \cdot V^*[\Lambda : \mathrm{I}] \leqq 2^{nf}$$

where

$$M_k = f(e_{1k}, \cdots, e_{nk}).$$

5. **Splitting. The number of reduced lattices is finite.** Up to now everything has worked for a division algebra of degree f over $\mathfrak{R}_0$ just as well as for a field $\mathfrak{J}$. Further progress depends on the structure of $\mathfrak{J}^*$. If $\mathfrak{J}$ is a field, then $\mathfrak{J}^*$ is isomorphic to the direct sum of a number of components $\mathfrak{R}$ and $\mathfrak{C}$. We first study this case.

The decomposition of $\mathfrak{J}^*$ is brought about by conjugation. One knows that $\mathfrak{J}/\mathfrak{R}_0$ has a determining number θ whose powers $1, \theta, \cdots, \theta^{f-1}$ constitute a basis for $\mathfrak{J}$. The number θ satisfies an irreducible equation in $\mathfrak{R}_0$ of degree f. Let θ^α and $\theta^\beta, \bar{\theta}^\beta$ (or in one row: $\theta^{(1)}, \cdots, \theta^{(f)}$) denote its r real and s pairs of complex conjugate roots. They define the f conjugations

$$\xi \to \xi^\alpha; \qquad \xi \to \xi^\beta, \qquad \xi \to \bar{\xi}^\beta$$

each of which projects $\mathfrak{J}$ isomorphically into $\mathfrak{R}$ or $\mathfrak{C}$. We use the notations

$$\xi^\alpha = x^\alpha, \qquad \xi^\beta = x_0^\beta + ix_1^\beta \qquad\qquad (\bar{\xi}^\beta = x_0^\beta - ix_1^\beta)$$

and call the $r+s$ numbers ξ^α, ξ^β the *splits*, and the f real numbers x^α; x_0^β, x_1^β

the *splitting coordinates* of ξ. The same applies to any element ξ^* of $\mathcal{J}^*$. The product $\zeta^* = \xi^*\eta^*$ has the splits

$$\zeta^\alpha = \xi^\alpha\eta^\alpha, \qquad \zeta^\beta = \xi^\beta\eta^\beta.$$

The arithmetician speaks of the different values of the indices α and β as the r real and s imaginary (*infinite prime*) *spots* of $\mathcal{J}$; for the sake of brevity we often drop the adjectives in parentheses. If a definite arrangement is desired, we write $\alpha = \alpha_1, \cdots, \alpha_r$; $\beta = \beta_1, \cdots, \beta_s$.

The splitting coordinates x^α; x_0^β, x_1^β are connected with the components $x_1, \cdots, x_f$ of ξ^* by the linear substitution

$$\Sigma = \|\sigma_1, \cdots, \sigma_f\| \tag{11}$$

where in the symbol on the right side each term stands for the column of its splitting coordinates (in a definite arrangement). The splitting of $\mathcal{J}^*$ into r components $\mathfrak{R}$ and s components $\mathfrak{C}$ is established as soon as it is certain that the absolute determinant

$$\Delta = \text{abs.}\,|\sigma_1, \cdots, \sigma_f|$$

of the matrix Σ is different from zero. For the particular basis $1, \theta, \cdots, \theta^{f-1}$ one sees that $(-2i)^s \cdot \Delta$ is the Vandermonde determinant of $\theta^{(1)}, \cdots, \theta^{(f)}$, and hence indeed $\Delta \neq 0$. This fact carries over to any basis $\sigma_1, \cdots, \sigma_f$ of $\mathcal{J}$.

The number in $\mathcal{J}^*$ with the splitting coordinates x^α; x_0^β, $-x_1^\beta$ is denoted by $\bar{\xi}^*$. As absolute value $|\xi^*|$ we introduce the greatest of the $r+s$ numbers $|\xi^\alpha|$, $|\xi^\beta|$. One could agree on other definitions, but this one is most convenient for our future applications. What usually is called a *unit* in $\mathcal{J}$ is a number of $\mathcal{J}$ which is a unity at all *finite* prime spots. None but the infinite prime spots matter for our investigation; hence we take the liberty of using the term "*unit*" for those numbers ϵ of $\mathcal{J}$ which are unities at all infinite prime spots, i.e., for which the $r+s$ equations $|\epsilon^\alpha| = 1$, $|\epsilon^\beta| = 1$ hold.

For any element δ^* of $\mathcal{J}^*$ one introduces the real matrix $\|d_{ab}\|$ of the linear operation $\xi^* \to \xi^*\delta^*$ in $\mathcal{J}^*$:

$$x_a \to \sum_b d_{ab}x_b \qquad \left(\sigma_a\delta^* = \sum_b d_{ba}\sigma_b\right)$$

and its characteristic polynomial

$$|te_{ab} - d_{ab}| = t^f - d_1t^{f-1} + \cdots \pm d_f.$$

d_1 and d_f are called trace (tr) and norm (Nm), respectively. In terms of the splitting coordinates our operation of multiplication splits into the transformations

$$x^\alpha \to x^\alpha d^\alpha; \qquad \xi^\beta \to \xi^\beta\delta^\beta,$$

each corresponding to a real or imaginary spot α or β. Of course, $\xi^\beta \to \xi^\beta \delta^\beta$ stands for

$$\overset{\beta}{x}_0 \to \overset{\beta}{x}_0 \overset{\beta}{d}_0 - \overset{\beta}{x}_1 \overset{\beta}{d}_1, \qquad \overset{\beta}{x}_1 \to \overset{\beta}{x}_0 \overset{\beta}{d}_1 + \overset{\beta}{x}_1 \overset{\beta}{d}_0.$$

Hence

$$\operatorname{tr}(\delta^*) = \sum_\alpha \overset{\alpha}{d} + 2\sum_\beta \overset{\beta}{d}_0,$$

$$\operatorname{Nm}(\delta^*) = \prod_\alpha \overset{\alpha}{d} \cdot \prod_\beta \{(\overset{\beta}{d}_0)^2 + (\overset{\beta}{d}_1)^2\}.$$

If $\delta^* = \delta$ is in $\mathcal{J}$ the d_{ab} are rational numbers. For a unit ϵ in $\mathcal{J}$ our formulas show that the determinant $\operatorname{Nm}(\epsilon)$ of the transformation $\xi^* \to \xi^* \epsilon$ is of absolute value 1 and hence as a rational number equal to ± 1.

Considering the trace $\operatorname{tr}(\xi^2)$ one readily verifies that $(2^b \Delta)^2$ is rational for any basis $\sigma_1, \cdots, \sigma_f$ and especially a rational integer for a basis of an order $[\mathcal{J}]$.

The transformation (4) in E^{nf},

$$\xi_i^* = \sum_k \eta_k^* \delta_{ik} \qquad (\delta_{ik} \text{ numbers in } \mathcal{J}) \tag{4}$$

splits into the components

$$\overset{\alpha}{\xi}_i = \sum_k \overset{\alpha}{\eta}_k \overset{\alpha}{\delta}_{ik}, \qquad \overset{\beta}{\xi}_i = \sum_k \overset{\beta}{\eta}_k \overset{\beta}{\delta}_{ik},$$

each α-component involving n, each β-component $2n$ real variables:

$$\overset{\alpha}{\xi}_k = \overset{\alpha}{x}_k; \qquad \overset{\beta}{\xi}_k = \overset{\beta}{x}_{k0} + i\overset{\beta}{x}_{k1}.$$

How closely can one approximate an element ξ^* of $\mathcal{J}^*$ by a number γ of our order $[\mathcal{J}]$ with the basis $\sigma_1, \cdots, \sigma_f$? For an appropriate γ in $[\mathcal{J}]$ the real components x_a' of $\xi^* - \gamma$,

$$\xi^* - \gamma = x_1'\sigma_1 + \cdots + x_f'\sigma_f,$$

will satisfy the inequalities $|x_a'| \leq \frac{1}{2}$, and thus

$$|\xi^* - \gamma| \leq \rho$$

where

$$\rho = \tfrac{1}{2} \cdot \max_{\alpha,\beta} \left(|\overset{\alpha}{\sigma}_1| + \cdots + |\overset{\alpha}{\sigma}_f|, \ |\overset{\beta}{\sigma}_1| + \cdots + |\overset{\beta}{\sigma}_f| \right).$$

The "circles" of radius ρ around all numbers γ of $[\mathcal{J}]$ cover the whole $\mathcal{J}^*$. (Such a radius was denoted by the letter r in R1, which now serves a different purpose.)

Let us now return to the situation explained at the end of the previous section and let V be the volume of K computed in terms of the splitting coordinates of $\xi_1^*, \cdots, \xi_n^*$. Then $V = V^*/\Delta^n$. Moreover we observe that

$f(\xi_1^*, \cdots, \xi_k^*, 0, \cdots, 0)$ is reduced with respect to the lattice $\Lambda^{(k)}$, and denoting by V_k the volume of the solid

$$f(\xi_1^*, \cdots, \xi_k^*, 0, \cdots, 0) < 1$$

in E^{kf} computed in terms of the splitting coordinates of $\xi_1^*, \cdots, \xi_k^*$, we obtain these fundamental inequalities for $M_k = f(e_{1k}, \cdots, e_{nk})$:

THEOREM I. *For a reduced* $f(\xi_1, \cdots, \xi_n)/\Lambda$ *one has*

$$(M_1 \cdots M_n)^f \cdot V[\Lambda:\mathrm{I}] \leqq (2^f\Delta)^n, \tag{12}$$

more generally

$$(M_1 \cdots M_k)^f \cdot V_k[\Lambda^{(k)}:\mathrm{I}^{(k)}] \leqq (2^f\Delta)^k. \tag{12$_k$}$$

At this point we introduce the further assumption:

(ii*) $f(\tau^*\mathfrak{x}^*) \leqq |\tau^*| \cdot f(\mathfrak{x}^*)$ (τ^* any element of $\mathfrak{J}^*$),

and henceforth the term "gauge function" is to be taken in this restricted sense. Following Minkowski's own argument, we then prove

THEOREM II. *For a reduced* $f(\xi_1, \cdots, \xi_n)/\Lambda$ *one always has*

$$j = [\Lambda:\mathrm{I}] \leqq (nf)!\left(\frac{4}{\pi}\right)^{ns} \cdot \left(\frac{\Delta}{f!}\right)^n \tag{13}$$

and more generally

$$j_k = [\Lambda^{(k)}:\mathrm{I}^{(k)}] \leqq (kf)!\left(\frac{4}{\pi}\right)^{ks}\left(\frac{\Delta}{f!}\right)^k \qquad (k = 1, \cdots, n). \tag{13$_k$}$$

Hence in any class of lattices belonging to the order $[\mathfrak{J}]$ there is always a lattice Λ which contains I and satisfies (13) and (13$_k$). Together with Lemma 2 this proves(6):

THEOREM III. *The number of classes of lattices belonging to a given order* $[\mathfrak{J}]$ *is finite.*

---

(6) This theorem is well known. We are concerned only with those lattices Λ over I which are in the class of $\mathfrak{L}$, but as our bounds (13) or the sharper bounds (35) depend on the order rather than on the special class it seemed worth while to mention Theorem III in passing. For a commutative field $\mathfrak{J}$ and its principal order $[\mathfrak{J}]$ E. Steinitz, Mathematische Annalen, vol. 71 (1912), pp. 328–354, and vol. 72 (1912), pp. 297–345, proved that the number for classes of any n is the same as for $n=1$, namely equal to the number of classes of ideals. See also I. Schur, Mathematische Annalen, vol. 71 (1912), pp. 355–367; W. Franz, Journal für die reine und angewandte Mathematik, vol. 171 (1934), pp. 149–161; C. Chevalley, *L'Arithmétique dans les Algèbres de Matrices*, Actualités Scientifiques et Industrielles, no. 323, 1936, in particular Theorems 3, 7 and 8.

(The proposition implies the corresponding one about classes of ideals.) Any gauge function will do for the proof, for instance

$$f(\xi_1^*, \cdots, \xi_n^*) = |\xi_1^*| + \cdots + |\xi_n^*|.$$

We shall soon see that much better upper bounds for the number of classes are obtained by using for f^2 the trace of a positive Hermitian form. However, our present Theorem II goes far beyond Theorem III because it deals with any gauge function f in conjunction with a lattice rather than with lattices alone.

Proof. Observe that the "octahedron"

$$|\xi_1^*| + \cdots + |\xi_n^*| < 1$$

contains no vector of Λ except the zero vector. Hence owing to Minkowski's chief inequality we find this upper bound for its volume W:

$$Wj \leqq (2/\Delta)^n.$$

Let $(\xi_1, \cdots, \xi_n)$ be a vector in Λ and ξ_k be the last nonvanishing one among its coordinates ξ_i. Then by the definition of reduction

$$f(\xi_1, \cdots, \xi_n) \geqq M_k = f(e_{1k}, \cdots, e_{nk}). \tag{14}$$

On the other hand the assumptions (iii) and (ii*) imply

$$\begin{aligned} f(\xi_1 e_1 + \cdots + \xi_n e_n) &\leqq M_1|\xi_1| + \cdots + M_n|\xi_n| \\ &= M_1|\xi_1| + \cdots + M_k|\xi_k|. \end{aligned} \tag{15}$$

Because of (8) the relations (14) and (15) are incompatible unless

$$|\xi_1| + \cdots + |\xi_n| = |\xi_1| + \cdots + |\xi_k| \geqq 1.$$

We base our computation of W upon the following general remark about gauge functions f in an n-dimensional vector space over $\mathfrak{R}$. If V is the volume of the gauge body $K: f(\mathfrak{x}) < 1$, then the integral $\int$ of e^{-f} over the whole space equals $n!V$. One simply evaluates the integral by decomposing the space into the infinitely thin shells

$$q \leqq f(\mathfrak{x}) < q + dq$$

and thus finds

$$\int = V \cdot \int_0^\infty e^{-q} \cdot nq^{n-1} dq = n!V.$$

Applying this remark to the gauge function $|\xi_1^*| + \cdots + |\xi_n^*|$ in our (nf)-dimensional vector space and to the gauge function $|\xi^*|$ in the f-dimensional space $\mathfrak{F}^*$, one gets this double value for $\int$:

$$(nf)!W = (f!w)^n,$$

w being the volume of the "cylinder" defined by

$$|\xi^*| < 1, \qquad \text{or by } |x^\alpha| < 1, \quad (x_0^\beta)^2 + (x_1^\beta)^2 < 1.$$

Therefore $w = 2^r\pi^s$.

(ii*) entails the property (ii$_0$) of §3, provided U is the group of units in our sense. From now on we shall abide by this convention and interpret the term "properly reduced" accordingly. Then the transformation $\xi^* \to \xi^* \cdot \epsilon$ (ϵ in U) and hence every special transformation (10) has the determinant ± 1 and thus the indices j_k for two lattices Λ and Λ' over I which are in the same family coincide: $j_k = j_k'$ for $k = 1, \cdots, n$.

The values γ^* of a *Hermitian form* in $\mathfrak{J}^*$,

$$\gamma^*(\mathfrak{x}^*) = \sum_{i,k} \xi_i^* \gamma_{ik}^* \bar{\xi}_k^* \qquad (\gamma_{ki}^* = \bar{\gamma}_{ik}^*) \tag{16}$$

are totally real in the sense that $\bar{\gamma}^* = \gamma^*$, or that even the β-splits $\gamma^\beta = g_0^\beta + ig_1^\beta = g^\beta$ of γ^* are real. What such a Hermitian form does is to associate a quadratic form $\{g_{ik}^\alpha\}$ with each real spot α and a Hermitian form $\{\gamma_{ik}^\beta\}$ with every imaginary spot β. The splits of $\gamma^*(\mathfrak{x}^*)$ are

$$g^\alpha = \sum_{i,k} x_i^\alpha g_{ik}^\alpha x_k^\alpha, \qquad g^\beta = \sum_{i,k} \xi_i^\beta \gamma_{ik}^\beta \bar{\xi}_k^\beta \tag{17}$$

where $x_i^\alpha = \xi_i^\alpha$ and ξ_i^β are the splits of ξ_i^*. The form $\gamma^*(\mathfrak{x}^*)$ is said to be positive if each of the r quadratic forms $\{g_{ik}^\alpha\}$ and each of the s Hermitian forms $\{\gamma_{ik}^\beta\}$ is positive definite.

We now apply our theory to the gauge function f introduced by

$$f^2 = \operatorname{tr}(\gamma^*(\mathfrak{x}^*)). \tag{18}$$

In terms of the splits (17) one has

$$f^2 = \sum_\alpha g^\alpha + 2\sum_\beta g^\beta. \tag{19}$$

The properties (i) to (iii) of §4 are readily verified; (ii*) is also fulfilled because of

$$f^2(\tau^*\mathfrak{x}^*) = \sum_\alpha |\tau^\alpha|^2 g^\alpha + 2\sum_\beta |\tau^\beta|^2 g^\beta.$$

6. Quaternion algebra of totally positive norm over a totally real field. Turning to noncommutative division algebras, we denote by $\mathfrak{Q}$ the quasi-field of quaternions

$$a = a_0 + a_1 i_1 + a_2 i_2 + a_3 i_3$$

whose components a_0, a_1, a_2, a_3 are arbitrary real numbers, and use the notations $\bar{a}$ and $|a|$ in the customary manner:

$$a\bar{a} = |a|^2 = a_0^2 + a_1^2 + a_2^2 + a_3^2.$$

For which of the noncommutative division algebras of finite degree over $\mathfrak{R}_0$ does the concept of infinite prime spots work in a way similar to that in the previous section for fields? I am going to describe one such situation without discussing the question whether or not it is the only one (though, as a matter of fact, it is).

Suppose we are given a field $\mathcal{E}$ of degree e over $\mathfrak{R}_0$ and two numbers ω_1, ω_2 in $\mathcal{E}$. We put $\omega_3 = \omega_1\omega_2$ and form the quaternion algebra $\mathcal{J}$ over $\mathcal{E}$ whose elements ξ are quadruples $(\xi_0, \xi_1, \xi_2, \xi_3)$ of numbers in $\mathcal{E}$,

$$\xi = \xi_0 + \xi_1\iota_1 + \xi_2\iota_2 + \xi_3\iota_3, \tag{20}$$

with this multiplication table for the unities $\iota_1, \iota_2, \iota_3$:

$$\iota_1^2 = -\omega_1, \qquad \iota_2^2 = -\omega_2, \qquad \iota_3^2 = -\omega_3;$$

$$\iota_1\iota_2 = -\iota_2\iota_1 = \iota_3, \qquad \iota_3\iota_1 = -\iota_1\iota_3 = \omega_1\iota_2, \qquad \iota_2\iota_3 = -\iota_3\iota_2 = \omega_2\iota_1.$$

The conjugate $\bar{\xi}$ is $\xi_0 - \xi_1\iota_1 - \xi_2\iota_2 - \xi_3\iota_3$ and

$$\xi\bar{\xi} = \xi_0^2 + \omega_1\xi_1^2 + \omega_2\xi_2^2 + \omega_3\xi_3^2.$$

If the equation

$$\xi_0^2 + \omega_1\xi_1^2 + \omega_2\xi_2^2 + \omega_3\xi_3^2 = 0 \tag{21}$$

has no solution $(\xi_0, \xi_1, \xi_2, \xi_3)$ in $\mathcal{E}$ except $(0, 0, 0, 0)$, then $\mathcal{J}$ is a division algebra of degree 4 over $\mathcal{E}$ and of degree $f = 4e$ over $\mathfrak{R}_0$. We assume $\mathcal{E}$ to be totally real (to have no imaginary infinite prime spot) and ω_1, ω_2 to be totally positive numbers in $\mathcal{E}$ (i.e., their e conjugates $\omega_1^\alpha, \omega_2^\alpha$ are all positive). Then the quadratic form of the variables $\xi_0, \xi_1, \xi_2, \xi_3$ at the left of (21) is positive definite in each conjugate $\mathcal{E}^\alpha$ of $\mathcal{E}$ and hence (21) has no solution except 0. Denoting as before by τ^α the conjugate of any number τ in $\mathcal{E}$ corresponding to the spot α of $\mathcal{E}$, we map (20) upon the element

$$\xi^\alpha = \xi_0^\alpha + \xi_1^\alpha(\omega_1^\alpha)^{1/2}\cdot i_1 + \xi_2^\alpha(\omega_2^\alpha)^{1/2}\cdot i_2 + \xi_3^\alpha(\omega_3^\alpha)^{1/2}\cdot i_3 \tag{22}$$

in $\mathfrak{Q}$. This "conjugation" is an isomorphic mapping and defines the "infinite quaternion prime spot" α of $\mathcal{J}$. (22) are the splits, and the $4e = f$ real numbers

$$x_0^\alpha = \xi_0^\alpha, \qquad x_1^\alpha = \xi_1^\alpha(\omega_1^\alpha)^{1/2}, \qquad x_2^\alpha = \xi_2^\alpha(\omega_2^\alpha)^{1/2}, \qquad x_3^\alpha = \xi_3^\alpha(\omega_3^\alpha)^{1/2}$$

are the "splitting coordinates," of ξ. Application to the elements ξ^*, equation (2*), of $\mathcal{J}^*$ is immediate.

There is only one thing to settle: The splitting coordinates $x_0^\alpha, x_1^\alpha, x_2^\alpha, x_3^\alpha$ arise from the components x_a by the substitution (11), each σ_a standing for

the column of its splitting coordinates. Is its determinant, whose absolute value will again be denoted by Δ, different from zero? To answer the question, let $(\tau_1, \cdots, \tau_e)$ be a basis of $\mathcal{E}$ and set $\Delta_0 = \text{abs.}|\tau_1, \cdots, \tau_e|$. From it we obtain the following basis of $\mathcal{J}$:

$$\tau_a, \quad \tau_a i_1, \quad \tau_a i_2, \quad \tau_a i_3 \qquad (a = 1, \cdots, e).$$

The Δ of this particular basis is given by

$$\Delta = \prod_a (\overset{a}{\omega_1}\overset{a}{\omega_2}) \cdot \Delta_0^4 = \text{Nm}\ \omega_3 \cdot \Delta_0^4.$$

Thus $\Delta \neq 0$ for this and consequently for any basis.

Incidentally Δ is a rational number for any basis of $\mathcal{J}$ and $4^e \cdot \Delta$ a rational integer if $\sigma_1, \cdots, \sigma_f$ is a basis of $[\mathcal{J}]$. The characteristic equation of the multiplication $\xi^* \to \xi^* \cdot \delta$ considered as a linear operation in $\mathcal{J}^*$ is the square of a polynomial (of degree $2e$), and so is the characteristic polynomial of the linear substitution (4) in E^{nf}.

The notion of unit and the absolute value $|\xi^*|$ of any element ξ^* of $\mathcal{J}^*$ are introduced as before. The constant on the right side of (13_k) is to be changed into

$$(kf)!\left(\frac{32}{\pi^2}\right)^{ek}\left(\frac{\Delta}{f!}\right)^k.$$

As gauge functions f we employ in particular those whose square equals

$$\tfrac{1}{4}\ \text{tr}\ (\gamma^*) = \sum_a g^a$$

where γ^* is any positive Hermitian form (16) in $\mathcal{J}^*$.

7. **The theorems of finiteness for quadratic forms.** After so many preliminaries which stake out the ground covered by our investigation, I now come to the core of the matter, which may be explained fairly completely by the simplest example $\mathcal{J} = \mathfrak{R}_0$. Here we have only one order $[\mathcal{J}]$ consisting of the ordinary integers $0, \pm 1, \pm 2, \cdots$ and only one class of lattices. For any given lattice Λ over I and any positive quadratic form

$$f^2(\mathfrak{x}) = \sum g_{ik}x_ix_k \qquad (g_{ki} = g_{ik})$$

the conditions of reduction read:

$$f^2(\mathfrak{x}) \geqq g_{kk} \qquad \text{whenever } \mathfrak{x} = (x_1, \cdots, x_n) \text{ is in } \Lambda_k.$$

Each of them is a linear inequality for the coefficients g_{ij}.

With the notation used in R1 we carry out Jacobi's transformation:

$$f^2 = q_1 z_1^2 + \cdots + q_n z_n^2.$$

The volume V of the ellipsoid $f^2 < 1$ is given by

$$V^2 = \omega_n^2/q_1 \cdots q_n,$$

ω_n being the volume of the n-dimensional sphere. Hence the inequality (12),

$$M_1 \cdots M_n V[\Lambda:\mathrm{I}] \leqq 2^n, \tag{23}$$

turns into

$$g_{11} \cdots g_{nn}[\Lambda:\mathrm{I}]^2 \leqq (2^n/\omega_n)^2 \cdot q_1 \cdots q_n. \tag{24}$$

As Minkowski observed, (23) may be proved much more easily for quadratic forms than for an arbitrary gauge function. By an argument similar to the one employed in proving Theorem II we see that the ellipsoid

$$f'^2 = \frac{q_1}{M_1^2} z_1^2 + \cdots + \frac{q_n}{M_n^2} z_n^2 < 1$$

contains no lattice vector except zero. Hence its volume V' satisfies the inequality

$$V'[\Lambda:\mathrm{I}] \leqq 2^n, \quad \text{and} \quad V' = M_1 \cdots M_n \cdot V.$$

If κ_n is a number such that the part of space covered by impenetrable n-dimensional spheres in any lattice arrangement may never exceed the proportion $\kappa_n : 1$ then we can even write $\kappa_n 2^n$ instead of 2^n and thus replace ω_n in (24) by $\pi_n = \omega_n/\kappa_n$. The most primitive choice is $\kappa_n = 1$; however, according to Blichfeldt's ingenious device(6),

$$\kappa_n = (n + 2) \cdot 2^{-1-n/2}$$

is a permissible and better value.

Making use of the inequalities

$$g_{ii} \geqq q_i$$

on the left side of (24) we get for the index $j = [\Lambda:\mathrm{I}]$ this upper bound

$$j \leqq 2^n/\pi_n \tag{25}$$

which is a considerable improvement over (13), $j \leqq n!$ For $n = 1, 2, 3$ it yields the result $j = 1$, to which the theory of reduction for binary and ternary forms owes its comparative simplicity.

For similar reasons

$$g_{11} \cdots g_{kk} \cdot j_k^2 \leqq (2^k/\pi_k)^2 \cdot q_1 \cdots q_k, \tag{24$_k$}$$

$$j_k \leqq 2^k/\pi_k \qquad (k = 1, \cdots, n). \tag{25$_k$}$$

Unless the lattice Λ satisfies the n inequalities (25$_k$) for its indices $j_k = [\Lambda^{(k)} : \mathrm{I}^{(k)}]$ there can be no Λ-reduced forms.

(6) H. F. Blichfeldt, Mathematische Annalen, vol. 101 (1929), pp. 605–608.

Dividing (24_k) by

$$g_{11} \cdots g_{k-1,k-1} \geqq q_1 \cdots q_{k-1}$$

we find that our reduced form satisfies the fundamental relations

$$q_k \geqq \lambda_k g_{kk} \tag{26}$$

where

$$\lambda_k = (j_k \pi_k / 2^k)^2. \tag{27}$$

This lower bound for q_k is much better than the corresponding one holding for the method of reduction studied in R1.

The *first theorem of finiteness* deals with the subset $\Lambda_k(=)$ of Λ_k to which a vector $\mathfrak{x}$ in Λ_k belongs if there exists a Λ-reduced positive quadratic form f^2 satisfying the *equation* $f^2(\mathfrak{x}) = g_{kk}$. *The set* $\Lambda_k(=)$ *is finite*. The proof is as in R1, but the upper bounds arrived at are a good deal lower. The first part of the proof yields the bounds

$$\lambda_i z_i^2 \leqq 1 \qquad (\text{for } i = k, k+1, \cdots, n)$$

where the λ_i are now defined by (27). In the second part one replaces the vector $\mathfrak{x}$ in $\Lambda_k(=)$ by $\mathfrak{x} - \mathfrak{x}_h$ where $\mathfrak{x}_h$ is any vector in $\Lambda^{(h)}$ ($h<k$) and observes that

$$f^2(\mathfrak{x} - \mathfrak{x}_h) \geqq g_{kk}.$$

This is true in particular if $\mathfrak{x}_h$ is in $\mathrm{I}^{(h)}$, and as in R1 one thus derives the relations

$$\lambda_h z_h^2 \leqq \rho^2 h \qquad (\rho = 1/2;\ h = 1, \cdots, k-1).$$

Once the discrete lattice Λ is given, the resulting universal upper bounds for $|x_n|, \cdots, |x_1|$ leave only a limited number of possibilities for a vector $\mathfrak{x} = (x_1, \cdots, x_n)$ in Λ.

The *second theorem of finiteness* shall be restated in a more natural and slightly more general form. Let $p \geqq 1$ and $w \geqq 0$ be given. With respect to the lattice Λ over I the positive quadratic form f^2 will be said to have the property $B(p, w)$ provided

$$f^2(\mathfrak{x}) \geqq \frac{1}{p} \cdot f^2(\mathfrak{e}_k) \tag{28}$$

for any vector $\mathfrak{x}$ in Λ_k, and

$$f^2(\mathfrak{e}_k - \mathfrak{x}_h) \geqq f^2(\mathfrak{e}_k) - w \cdot f^2(\mathfrak{e}_h) \tag{28'}$$

for $h<k$ and any vector $\mathfrak{x}_h$ in $\Lambda^{(h)}$. Again, each of these conditions is a linear inequality for the coefficients g_{ij} of f^2. We maintain:

Given two lattices Λ *and* Λ' *over* I, *there is only a limited number of linear*

transformations carrying Λ into Λ' and at the same time capable of carrying an unspecified Λ-reduced form f^2 into an unspecified f'^2 which has the property $B(p, w)$ with respect to Λ'.

We write the transformation as

$$\mathfrak{x} = \sum_i x_i \mathfrak{e}_i = \sum_i y_i \mathfrak{d}_i:$$

if $(x_1, \cdots, x_n)$ is in Λ, then $(y_1, \cdots, y_n)$ is in Λ', and *vice versa*. In particular, the $\mathfrak{d}_1, \cdots, \mathfrak{d}_n$ are vectors in Λ. (p, $\mathfrak{e}_i$, $\mathfrak{d}_i$ were denoted in R1 by ρ^2, $\mathfrak{b}_i$, $\mathfrak{s}_i$.) More explicitly as has been done in R1, we divide the row of indices $1, \cdots, n$ into a number of sections by means of the subspaces

$$E_k = [\mathfrak{e}_1, \cdots, \mathfrak{e}_k], \qquad E_k' = [\mathfrak{d}_1, \cdots, \mathfrak{d}_k] \qquad (k = 0, 1, \cdots, n).$$

We pick out those $k = l_0, l_1, \cdots, l_v$,

$$0 = l_0 < l_1 < \cdots < l_{v-1} < l_v = n$$

for which $E_k = E_k'$, and divide the range of k into the v sections

$$l_{u-1} < k \leqq l_u \qquad (u = 1, \cdots, v).$$

We then study the possibilities for transformations $(\mathfrak{d}_1, \cdots, \mathfrak{d}_n)$ with given $l_1, \cdots, l_{v-1}$.

By the analogues of Theorems 8_p and 9_p we have

$$g'_{kk} \leqq p g_{kk} \qquad (k = 1, \cdots, n) \tag{29}$$

and moreover

$$g_{i+1,i+1} \leqq p g_{ii} \tag{30}$$

whenever i and $i+1$ are in the same section. Consider a $\mathfrak{d}_k$ of the last section $(l_{v-1} < k \leqq n)$. The first part of the proof in R1 yields for $\mathfrak{x} = \mathfrak{d}_k$ the simple upper bounds

$$\lambda_h z_h^2 \leqq p^{\{k-h\}+1}$$

if h also belongs to the last section, $\{k\}$ denoting 0 or k according as $k \leqq 0$ or $k > 0$. The second part requires a slight modification. Suppose h lies in the uth section $(u < v)$, and set for the moment $l_u = l$. Since $E_l = E_l'$, the vectors in $\Lambda^{(l)}$ are obtained from the expression $y_1\mathfrak{d}_1 + \cdots + y_l\mathfrak{d}_l$ by running $(y_1, \cdots, y_l, 0, \cdots, 0)$ over $\Lambda'^{(l)}$. Hence, according to the postulate (28′):

$$f'^2(\mathfrak{e}_k - \mathfrak{y}') \geqq g'_{kk} - w g'_{ll}$$

for any vector $\mathfrak{y}'$ in $\Lambda'^{(l)}$, or

$$f^2(\mathfrak{d}_k - \mathfrak{x}') \geqq g'_{kk} - w g'_{ll}$$

for any vector $\mathfrak{x}'$ in $\Lambda^{(l)}$, *a fortiori* for any vector in $\Lambda^{(h)}$, *a fortiori* for any vector $\mathfrak{x}'$ in $I^{(h)}$. Following the same argument as in R1, one gets the inequality

$$wg_{ll}' + \rho^2 hg_{hh} \geqq \lambda_h g_{hh} z_h^2 \qquad (\rho = 1/2).$$

But because h and l are in the same section, (30) and (29) lead to

$$g_{ll} \leqq p^{l-h} g_{hh}, \qquad g_{ll}' \leqq pg_{ll} \leqq p^{l-h+1} g_{hh}$$

and thus finally

$$\lambda_h z_h^2 \leqq h\rho^2 + w\cdot p^{l_u - h+1} \qquad (l_{u-1} < h \leqq l_u;\ u = 1, \cdots, v-1).$$

It is clear how the same procedure applies to a k in the lower sections. Denoting the values of the variables $z_1, \cdots, z_n$ for $\mathfrak{x} = \mathfrak{b}_k$ by $z_{1k}, \cdots, z_{nk}$,one finds:

$z_{hk} = 0$ if h is in a higher section than k;

$\lambda_h z_{hk}^2 \leqq p^{|k-h|+1}$ if h and k are in the same section;

$\lambda_h z_{hk}^2 \leqq h\rho^2 + w\cdot p^{l-h+1}$ if h is in a lower section than k which ends with l.

8. **Modifications in arbitrary fields and quasi-fields.** Our next concern is to examine whether any serious modifications of the procedure just described arise in the two general cases of a field and a quaternion algebra over a field. Take the case of the field first. With a positive Hermitian form γ^* in $\mathcal{J}^*$ we combine its trace f^2:

$$(31)\qquad f^2(\xi_1^*, \cdots, \xi_n^*) = \sum_\alpha \sum_{i,k} \overset{\alpha}{g}_{ik} \overset{\alpha}{x}_i \overset{\alpha}{x}_k + 2\sum_\beta \sum_{i,k} \overset{\beta}{\gamma}_{ik} \overset{\beta}{\xi}_i \overset{\beta}{\bar{\xi}}_k \qquad (\overset{\beta}{\xi}_k = \overset{\beta}{x}_{k0} + i\overset{\beta}{x}_{k1}).$$

γ^* is called reduced with respect to Λ if the gauge function f is, i.e., if

$$(32)\qquad f^2(\xi_1, \cdots, \xi_n) \geqq \operatorname{tr}(\gamma_{kk}^*) = M_k^2$$

for any vector $(\xi_1, \cdots, \xi_n)$ in Λ_k. Each part is subjected to its Jacobi transformation:

$$\sum_{i,k} \overset{\alpha}{g}_{ik} \overset{\alpha}{x}_i \overset{\alpha}{x}_k = \sum_i \overset{\alpha}{q}_i (\overset{\alpha}{z}_i)^2,$$
$$\sum_{i,k} \overset{\beta}{\gamma}_{ik} \overset{\beta}{\xi}_i \overset{\beta}{\bar{\xi}}_k = \sum_i \overset{\beta}{q}_i |\overset{\beta}{\zeta}_i|^2.$$

Besides

$$\operatorname{tr}(q_i) = \sum_\alpha \overset{\alpha}{q}_i + 2\sum_\beta \overset{\beta}{q}_i, \qquad \operatorname{Nm}(q_i) = \prod_\alpha \overset{\alpha}{q}_i \cdot \prod_\beta (\overset{\beta}{q}_i)^2$$

we introduce the mean value $\langle q_i \rangle$ by

$$f\cdot\langle q_i\rangle = \operatorname{tr}(q_i).$$

In terms of the coordinates x_k^α; x_{k0}^β, x_{k1}^β the volume V of the ellipsoid $f^2(\mathfrak{x}^*)<1$ is

$$\omega_{nf} \text{ divided by } 2^{ns}\left(\prod_i \mathrm{Nm}\, q_i\right)^{1/2}.$$

Instead of applying Minkowski's second inequality to the present gauge function we again consider the ellipsoid

$$f'^2(\mathfrak{x}^*) = \mathrm{tr}\left(\sum_i \frac{q_i}{M_i^2}\zeta_i\bar{\zeta}_i\right) < 1$$

which contains no lattice vector except zero, and thus establish the inequality

$$\prod_i (\mathrm{tr}\,\overset{*}{\gamma}_{ii})^f\cdot[\Lambda:\mathrm{I}]^2 \leqq \frac{(4^{r+s}\Delta^2)^n}{\pi_{nf}^2}\cdot\prod_i \mathrm{Nm}\, q_i \tag{33}$$

for any reduced γ^*/Λ.

Now enters the only new feature: Making use of the inequality between arithmetic and geometric means in the form

$$\mathrm{Nm}\, q_i \leqq \langle q_i\rangle^f$$

we infer from

$$\overset{\alpha}{q}_i \leqq \overset{\alpha}{g}_{ii}, \qquad \overset{\beta}{q}_i \leqq \overset{\beta}{\gamma}_{ii}$$

the relation

$$\langle\overset{*}{\gamma}_{ii}\rangle^f \geqq \langle q_i\rangle^f \geqq \mathrm{Nm}\, q_i,$$

and then (33) yields the following upper bound for $j=[\Lambda:\mathrm{I}]$:

$$j \leqq 1/\mu_n \quad \text{with the abbreviation} \quad \mu_n = \frac{\pi_{nf}\cdot f^{nf/2}}{(2^{r+s}\Delta)^n}.$$

For the same reasons

$$\prod_{i=1}^k \langle\overset{*}{\gamma}_{ii}\rangle^f(\mu_k j_k)^2 \leqq \prod_{i=1}^k \mathrm{Nm}\, q_i \tag{34}$$

and hence

$$\mu_k j_k \leqq 1. \tag{35}$$

These estimates are an improved substitute for Theorem II. Combining (34) with

$$\prod_{i=1}^{k-1}\langle\overset{*}{\gamma}_{ii}\rangle^f \geqq \prod_{i=1}^{k-1}\langle q_i\rangle^f \geqq \prod_{i=1}^{k-1}\mathrm{Nm}\, q_i$$

one gets

$$\text{Nm } q_k \geqq (\mu_k j_k)^2 \cdot \langle \gamma_{kk}^* \rangle^f. \tag{36}$$

Not only does this inequality establish a lower bound for the trace of q_k,

$$\text{tr } q_k \geqq (\mu_k j_k)^{2/f} \cdot \text{tr } \gamma_{kk}^*,$$

but it also shows that the geometric mean of the conjugates of q_k is not much smaller than their arithmetic mean. Therefore none of the conjugates can be much smaller than their arithmetic mean. We have a special case of the situation dealt with by the following

LEMMA 5. *Let $f_1, \cdots, f_m$ be positive integers and $u_1, \cdots, u_m$; $v_1, \cdots, v_m$ two rows of positive numbers. We set*

$$f = f_1 + \cdots + f_m,$$
$$f \cdot \langle u \rangle = f_1 u_1 + \cdots + f_m u_m,$$
$$\text{Nm } u = u_1^{f_1} \cdots u_m^{f_m}.$$

If $u_\alpha \leqq v_\alpha$ $(\alpha = 1, \cdots, m)$ and

$$\text{Nm } u \geqq \mu \cdot \langle v \rangle^f \tag{37}$$

with some constant $\mu \geqq 1$, then

$$u_\alpha \geqq \lambda_\alpha \cdot \langle v \rangle$$

where λ_α depends on μ but not on the u and v.

(In our case r among the weights f_α are 1 and s of them equal 2.)

Proof. In the trivial case $m = 1$ one determines λ by

$$\lambda^f = \mu. \tag{38}$$

If $m > 1$ we set $u_1 = \lambda \cdot \langle v \rangle$ and assume $\lambda \leqq 1$. Then

$$\text{Nm } u = \lambda^{f_1} \langle v \rangle^{f_1} \cdot u_2^{f_2} \cdots u_m^{f_m} \leqq \lambda^{f_1} \langle v \rangle^{f_1} \langle u \rangle_1^{f - f_1}.$$

Here $\langle u \rangle_1$ denotes the arithmetic mean of $u_2, \cdots, u_m$ formed with the weights $f_2, \cdots, f_m$ of sum $f - f_1$:

$$\begin{aligned}(f - f_1) \cdot \langle u \rangle_1 &= f_2 u_2 + \cdots + f_m u_m = f \cdot \langle u \rangle - f_1 u_1 \\ &= f \cdot \langle u \rangle - f_1 \lambda \langle v \rangle \leqq (f - f_1 \lambda) \langle v \rangle.\end{aligned}$$

Therefore

$$\text{Nm } u \leqq \mu \cdot \langle v \rangle^f$$

where

$$\mu = \lambda^{f_1}\left(\frac{f - f_1\lambda}{f - f_1}\right)^{f-f_1}. \tag{39}$$

As its logarithmic derivative shows,

$$\frac{d\mu}{\mu} = \frac{f_1(1-\lambda)}{f - f_1\lambda}\cdot\frac{fd\lambda}{\lambda}; \tag{40}$$

this function $\mu(\lambda)$ maps the interval $0 \leqq \lambda \leqq 1$ monotonically upon $0 \leqq \mu \leqq 1$ and thus will assume the given value μ $(\leqq 1)$ for a certain $\lambda = \lambda_1$ $(\leqq 1)$. Thus we cannot have $u_1 < \lambda_1 \cdot \langle v \rangle$ under the condition (37).

(We wish to obtain the best value for the constant λ_1. If one is content with a little less, one may choose λ_1 according to the equation

$$\lambda_1^{f_1}\cdot\left(\frac{f}{f - f_1}\right)^{f-f_1} = \mu,$$

or even, as

$$\left(1 + \frac{f_1}{f - f_1}\right)^{(f-f_1)/f_1} < e \qquad (= \text{basis of natural logarithms}),$$

$$(\lambda_1 e)^{f_1} = \mu.$$

Incidentally, formula (40) holds good also for the function (38) which rules the trivial case $m = 1, f_1 = f$.)

In this way we ascertain constants λ_k, λ_k' such that

$$\text{each} \quad \overset{\alpha}{q_k} \geqq \lambda_k \cdot \langle \overset{*}{\gamma}_{kk} \rangle \quad \text{and each} \quad \overset{\beta}{q_k} \geqq \lambda_k' \cdot \langle \overset{*}{\gamma}_{kk} \rangle.$$

In case there is only one infinite prime spot, λ_k and λ_k' are determined by the relation

$$(\mu_k j_k)^2 = \lambda_k = \lambda_k'^2$$

in case of several spots by the equations

$$(\mu_k j_k)^2 = \lambda_k\left(\frac{f - \lambda_k}{f - 1}\right)^{f-1} = \lambda_k'^2\left(\frac{f - 2\lambda_k'}{f - 2}\right)^{f-2}$$

together with $\lambda_k \leqq 1$ and $\lambda_k' \leqq 1$.

Similarly for the other case studied in §6, that of a quaternion algebra $\mathfrak{J}$ with totally positive relative norm over a totally real field. The constants μ_k, λ_k in the inequalities

$$\mu_k j_k \leqq 1 \quad \text{and} \quad \overset{a}{q_k} \geqq \lambda_k \cdot \langle \overset{*}{\gamma}_{kk} \rangle$$

are then given by

$$\mu_k = \pi_{4k_e} \Delta^{-k}(e/4)^{2ke},$$

$$(\mu_k j_k)^{1/2} = \lambda_k \quad \text{or} \quad \lambda_k\left(\frac{e-\lambda_k}{e-1}\right)^{e-1} \qquad (\lambda_k \leqq 1),$$

according as $e = \frac{1}{2}f$ is 1 or is greater than 1.

After this the proofs for the first and second theorems of finiteness roll along as before.

9. **The pattern of equivalent cells.** The Hermitian forms $\{\gamma_{ik}^*\}$ constitute a linear space of

$$N = f\cdot\tfrac{1}{2}n(n-1) + (r+s)n = f\cdot\tfrac{1}{2}n(n+1) - sn \qquad (\text{field } \mathfrak{J})$$

or

$$N = en(2n-1) \qquad (\text{quasi-field } \mathfrak{J})$$

dimensions, the positive ones a convex cone G in that space. G is an open set; we operate within G throughout. "Form" means any positive Hermitian form.

Let Λ be a lattice over I. A Λ-reduced form γ^* has been characterized by the inequalities

$$(32) \qquad f^2(\xi_1, \cdots, \xi_n) \geqq f^2(e_{1k}, \cdots, e_{nk})$$

holding for $f^2 = \text{tr}\ (\gamma^*)$ whenever $(\xi_1, \cdots, \xi_n)$ is in Λ_k. For a given vector $(\xi_1, \cdots, \xi_n)$ the equality sign in (32) will hold identically for all Hermitian forms γ^* only if

$$(\xi_1, \cdots, \xi_n) = \epsilon\cdot(e_{1k}, \cdots, e_{nk}) \qquad (\epsilon \text{ a unit}),$$

as follows at once from the expression (31). For any other vector $(\xi_1, \cdots, \xi_n)$ the equation determines a $(N-1)$-dimensional hyperplane in our N-dimensional linear space of forms. This remark justifies our definition of "properly reduced" in terms of the group U of units.

The forms γ^* which are reduced with respect to Λ make up a convex cone G_Λ in G. The properly reduced forms are the inner points of G_Λ; see R1, p. 150. G_Λ may be empty; indeed it will be so unless the indices j_k of Λ satisfy the inequalities (35). Even if it is not empty it may be without inner points. Theorem 10 in R1, together with the first theorem of finiteness, proves:

THEOREM IV. *If G_Λ has inner points, then G_Λ is a convex pyramid defined within G by a limited number of linear inequalities.*

A linear mapping $\mathfrak{x}\to\mathfrak{x}'$ of $E^n/\mathfrak{J}$ upon itself is one satisfying the conditions $(\mathfrak{x}_1+\mathfrak{x}_2)' = \mathfrak{x}_1' + \mathfrak{x}_2'$ and $(\delta\cdot\mathfrak{x})' = \delta\cdot\mathfrak{x}'$ for any number δ in $\mathfrak{J}$. We also require that $\mathfrak{x}=0$ is the only vector whose image $\mathfrak{x}'$ equals 0. If $\mathfrak{d}_1, \cdots, \mathfrak{d}_n$ is any basis of $E^n/\mathfrak{J}$ the mapping S may be defined by giving the images $\mathfrak{d}_i' = \mathfrak{d}_i S$ of the $\mathfrak{d}_i$. The mapping S carries a form γ^* into a form γ_S^* according to the equation $\gamma_S^*(\mathfrak{x}S) = \gamma^*(\mathfrak{x})$. An order $[\mathfrak{J}]$ in $\mathfrak{J}$ and a lattice $\mathfrak{L}$ belonging to the order $[\mathfrak{J}]$ are supposed to be given. The linear mappings S which leave $\mathfrak{L}$ invariant are

said to form the *modular group*(7). In terms of a basis $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$ of $E^n/\mathfrak{J}$ the lattice $\mathfrak{L}$ is represented by Λ:

$$\mathfrak{x} = \eta_1\mathfrak{b}_1 + \cdots + \eta_n\mathfrak{b}_n \quad \text{in} \quad \mathfrak{L} \cdot \rightleftarrows \cdot (\eta_1, \cdots, \eta_n) \quad \text{in} \quad \Lambda,$$

and a form $\gamma^*(\mathfrak{x})$ is represented by a form $\Gamma^*(\eta_1, \cdots, \eta_n)$:

$$\gamma^*(\eta_1\mathfrak{b}_1 + \cdots + \eta_n\mathfrak{b}_n) = \Gamma^*(\eta_1, \cdots, \eta_n). \tag{41}$$

The linear mapping defined by $\mathfrak{b}_i \rightarrow \mathfrak{b}_i'$ carries $\eta_1\mathfrak{b}_1 + \cdots + \eta_n\mathfrak{b}_n$ into $\eta_1\mathfrak{b}_1' + \cdots + \eta_n\mathfrak{b}_n'$; hence it leaves $\mathfrak{L}$ invariant and thus belongs to the modular group if and only if $\Lambda = \Lambda'$, Λ and Λ' being the representations of $\mathfrak{L}$ in terms of $\mathfrak{b}_i$ and $\mathfrak{b}_i'$. For a vector $\mathfrak{b}$ in $\mathfrak{L}$ there are not more than a finite number of units ϵ such that $\epsilon\mathfrak{b}$ also is in $\mathfrak{L}$. Indeed, for the splits of $\mathfrak{x} = \epsilon\mathfrak{b}$ one finds

$$|\xi^\alpha| = |\delta^\alpha|, \quad |\xi^\beta| = |\delta^\beta|, \quad \text{a fortiori} \quad |\xi^\alpha| \leqq |\delta^\alpha|, \quad |\xi^\beta| \leqq |\delta^\beta|,$$

which in view of the discrete nature of $\mathfrak{L}$ proves the point.

We want to divide G without gaps and overlappings into domains which are mutually equivalent under the modular group. We shall introduce these cells first as entities which have nothing to do with Hermitian forms, adopting a criterion of identity other than the set-theoretic one. The systematic place for this introduction would have been at the end of §3. Only afterwards shall we explain the meaning of the phrase "a form lies in a cell." Here are the definitions:

A semi-basis $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$ of $\mathfrak{L}$ determines a cell $Z(\mathfrak{b}_1, \cdots, \mathfrak{b}_n)$; the semi-basis $\bar{\mathfrak{b}}_1, \cdots, \bar{\mathfrak{b}}_n$ is said to determine the same cell if

$$\bar{\mathfrak{b}}_i = \epsilon_i\mathfrak{b}_i \qquad (\epsilon_i \text{ units}). \tag{42}$$

Let S be an operation of the modular group. The image Z_S of $Z = Z(\mathfrak{b}_1, \cdots, \mathfrak{b}_n)$ is defined as $Z(\mathfrak{b}_1', \cdots, \mathfrak{b}_n')$ where $\mathfrak{b}_i' = \mathfrak{b}_iS$. (Notice that if Z is written as $Z(\bar{\mathfrak{b}}_1, \cdots, \bar{\mathfrak{b}}_n)$, $\bar{\mathfrak{b}}_i = \epsilon_i\mathfrak{b}_i$, then $Z(\bar{\mathfrak{b}}_1', \cdots, \bar{\mathfrak{b}}_n')$ is the same Z_S because $\bar{\mathfrak{b}}_i' = \epsilon_i\mathfrak{b}_i'$; thus Z_S is independent of the fixation of the defining semi-basis $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$.) Those S of the modular group for which $Z_S = Z$ shall be denoted by J_Z; they form a finite group $\{J_Z\}$. Indeed, for such an $S = J_Z$ one must have

$$\mathfrak{b}_i' = \mathfrak{b}_iS = \sigma_i\mathfrak{b}_i \qquad (\sigma_i \text{ a unit}), \tag{43}$$

and the J_Z are those mappings of the special form (43) which leave $\mathfrak{L}$ invariant. (In terms of another defining semi-basis $\bar{\mathfrak{b}}_i = \epsilon_i\mathfrak{b}_i$ the same J_Z is expressed by $\bar{\mathfrak{b}}_i' = \epsilon_i\sigma_i\epsilon_i^{-1}\cdot\bar{\mathfrak{b}}_i$.) Any operation S of the modular group has the same effect upon Z as J_ZS.

---

(7) If one feels that this term ought to be reserved for the group which is fundamental in the theory of the modules of the theta functions then a new word, say "lattice group," is indicated for our purpose.

In terms of $(\mathfrak{b}_1, \cdots, \mathfrak{b}_n)$ the lattice $\mathfrak{L}$ is represented by an admissible lattice Λ, i.e., by a lattice Λ over I which is equivalent to $\mathfrak{L}$. Hence to the cell $Z=Z(\mathfrak{b}_1, \cdots, \mathfrak{b}_n)$ there corresponds a family of admissible lattices Λ, and the same family to each equivalent cell Z_S. We have a one-to-one correspondence between the classes of equivalent cells on the one hand, and the families of admissible lattices Λ on the other. We distinguish them by different colors. The operations J_Z are represented by the operations J_Λ in terms of the basis $(\mathfrak{b}_1, \cdots, \mathfrak{b}_n)$.

Now we come to the realization of cells as point sets in G. A form γ^* is said to lie in $Z=Z(\mathfrak{b}_1, \cdots, \mathfrak{b}_n)$ if $(\mathfrak{b}_1, \cdots, \mathfrak{b}_n)$ is reduced with respect to γ^*, i.e., if for $f^2=\mathrm{tr}(\gamma^*)$ one has $f^2(\mathfrak{x}) \geqq f^2(\mathfrak{b}_k)$ whenever $\mathfrak{x}$ is in $\mathfrak{L}$ and outside $[\mathfrak{b}_1, \cdots, \mathfrak{b}_{k-1}]$. Because (42) implies

$$f^2(\mathfrak{b}_k) = f^2(\mathfrak{b}_k), \qquad [\mathfrak{b}_1, \cdots, \mathfrak{b}_{k-1}] = [\mathfrak{b}_1, \cdots, \mathfrak{b}_{k-1}],$$

the definition is independent of the fixation of the defining semi-basis $\mathfrak{b}_1, \cdots, \mathfrak{b}_n$. If γ^* lies in $Z=Z(\mathfrak{b}_1, \cdots, \mathfrak{b}_n)$ then the transform Γ^* introduced by (41) lies in G_Λ, and γ_S^* lies in Z_S.

The fact that there always exists a reduced semi-basis for a given γ^* and the concluding sentence of §3 can now be stated thus:

(a) Every point γ^* lies in at least one cell Z.

(b) An inner point of a cell Z cannot lie in a cell Z' unless Z' is the same as Z (or briefly: different cells have no inner points in common).

The fact (a) will of course not be altered by suppressing all *empty* cells and their colors. Thus we have to look only for those admissible Λ whose indices satisfy the conditions (35); and this brings the colors down to a limited number. Will (a) still prevail after suppressing all cells without inner points and their colors? The answer is affirmative because there is no inner clustering of cells in G. This is a consequence of the second theorem of finiteness, which now takes on the following form. Let $\mathfrak{a}_1, \cdots, \mathfrak{a}_n$ be a semi-basis of $\mathfrak{L}$, $p \geqq 1$ and $w \geqq 0$. The form γ^* is said to lie in $Z(\mathfrak{a}_1, \cdots, \mathfrak{a}_n | p, w)$ if

$$f^2(\mathfrak{x}) \geqq \frac{1}{p} \cdot f^2(\mathfrak{a}_k)$$

whenever $\mathfrak{x}$ is in $\mathfrak{L}$ and outside $[\mathfrak{a}_1, \cdots, \mathfrak{a}_{k-1}]$, and if, moreover,

$$f^2(\mathfrak{a}_k - \mathfrak{x}_h) \geqq f^2(\mathfrak{a}_k) - w \cdot f^2(\mathfrak{a}_h)$$

whenever $h<k$ and $\mathfrak{x}_h$ is in $\mathfrak{L}$ and $[\mathfrak{a}_1, \cdots, \mathfrak{a}_h]$.

THEOREM V. *There is only a finite number of operations S of the modular group such that the image Z_S into which a given cell Z is thrown by S will have points in common with the domain $Z(\mathfrak{a}_1, \cdots, \mathfrak{a}_n | p, w)$.*

Application to $p=1$, $w=0$ proves in particular that a cell borders on not more than a finite number of other cells. And since $Z(\mathfrak{a}_1, \cdots, \mathfrak{a}_n | p, w)$ sweeps

over the whole G if p and w increase to infinity we are sure that the cells cluster around no point in the interior of G ("the modular group is properly discontinuous in G"). We therefore definitely admit only those colors whose cells are N-dimensional solids, i.e., have inner points. In our summary we talk of them as point sets in G.

THEOREM VI. (a) *G is divided into a pattern of cells, each cell bearing a color out of a finite palette of colors. The cells cover G without gaps and overlappings. Each cell is a solid convex pyramid (in G). The mappings of the modular group leave this design, including its coloring, invariant. Any two cells of the same color can be carried one into the other by an operation of the modular group.*

(b) *Given a point in G and a cell Z one can assign a neighborhood $\mathfrak{N}$ to the point such that there is only a limited number of operations S of the modular group for which the image Z_S penetrates into $\mathfrak{N}$.*

(c) *The operations of the modular group which carry a cell into itself form a finite subgroup. This group of linear operations in the vector space $E^n/\mathfrak{J}$ is equivalent (in $\mathfrak{J}$) to a group whose elements are of the special form*

$$\xi_1 \to \xi_1\epsilon_1, \cdots, \xi_n \to \xi_n\epsilon_n \qquad (\epsilon_i \text{ units}).$$

(Of course, in view of statement (c) the statement (b) could have been replaced by the simpler one that only a finite number of cells penetrate into $\mathfrak{N}$.)

We form a *nucleus* by selecting one cell Z_c of each color c. All cells adjacent to the nuclear cells form a *wreath* around the nucleus. Here the word "adjacent" may be interpreted either in the wide sense of "having a point in common," or in the narrower sense of "having a wall of $N-1$ dimensions in common."

THEOREM VII. *Determine for each cell $Z_c^!$ of color c in the wreath an operation $S_c^!$ of the modular group which maps the nuclear cell Z_c of color c into $Z_c^!$. The $S_c^!$ thus selected, together with the operations J_c of the modular group which carry Z_c into itself, generate the whole group if all colors c are taken into account.*

Were it not for the groups $\{J_\Lambda\}$ the nucleus would form a fundamental domain. As it is, one has first to replace in our construction each G_Λ by a part G_Λ' which in G_Λ is a fundamental domain for the finite group of special transformations

$$J_\Lambda: \quad \xi_k \to \xi_k\epsilon_k \qquad (\epsilon_k \text{ a unit})$$

carrying G_Λ into itself. The effect of J_Λ upon the coefficient $\overset{*}{\gamma}_{ik}$ is described by

$$\overset{*}{\gamma}_{ik} \to \epsilon_i\overset{*}{\gamma}_{ik}\bar{\epsilon}_k.$$

If in one split α the transformation of the variable $\overset{\alpha}{\gamma}_{ik} = \Xi$,

$$\Xi \to \overset{\alpha}{\epsilon}_i \Xi \overset{\alpha}{\bar{\epsilon}}_k = \overset{\alpha}{\epsilon}_i \Xi (\overset{\alpha}{\epsilon}_k)^{-1},$$

is the identity, then the same is true of every split. Hence it is sufficient to consider one split α only, and after choosing it we write simply $\gamma_{ik}^{\alpha}=\gamma_{ik}$, $\epsilon_i^{\alpha}=\epsilon_i$. If the transformation $\Xi\rightarrow\epsilon_1\Xi\epsilon_2^{-1}$ is the identity, one must have $\epsilon_1=\epsilon_2$ as the specialization $\Xi=1$ shows, and $\Xi\rightarrow\epsilon_2\Xi\epsilon_1^{-1}$ is also the identity. Moreover, if $\Xi\rightarrow\epsilon_1\Xi\epsilon_2^{-1}$ and $\Xi\rightarrow\epsilon_2\Xi\epsilon_3^{-1}$ are identities, then $\Xi\rightarrow\epsilon_1\Xi\epsilon_3^{-1}$ is. Consequently we may well limit ourselves first to the coefficients γ_{ik} $(i<k)$ on one side of the diagonal, and then more particularly to

$$\Xi_1 = \gamma_{12},\ \Xi_2 = \gamma_{23},\ \cdots,\ \Xi_{n-1} = \gamma_{n-1,n}.$$

Let us at once consider the most disagreeable case, that of a quaternion quasi-field $\mathcal{J}$ as described in §6.

The group $\{J_{\mathrm{A}}\}$ induces a group of transformations of the type

$$\Xi \rightarrow \epsilon_1 \Xi \bar{\epsilon}_2 \qquad (|\epsilon_1| = |\epsilon_2| = 1)$$

for $\gamma_{12}=\Xi_1=\Xi$. This is a finite group of orthogonal transformations J in the space of the four components X_0, X_1, X_2, X_3 of the variable quaternion Ξ. Denote by ΞJ the transform of Ξ by J. The simplest way of ascertaining a fundamental domain for this group $\{J\}$ is as follows: One chooses a point $\Xi=\mathrm{A}$ which differs from all its transforms $\mathrm{A}J$ ($J\neq$identity). The fundamental domain consists of all points Ξ which are nearer to the center A than to the other equivalent centers $\mathrm{A}J$ and is thus characterized by the inequalities

$$\Xi\cdot\overline{\mathrm{A} - \mathrm{A}J} + (\mathrm{A} - \mathrm{A}J)\cdot\bar{\Xi} \geqq 0.$$

Fortunately these are linear inequalities, namely of the form

$$a_0X_0 + a_1X_1 + a_2X_2 + a_3X_3 \geqq 0$$

(a_0, a_1, a_2, a_3 being the components of $\mathrm{A}-\mathrm{A}J$). After having done this we limit ourselves to those operations J_{A} which leave Ξ_1 unchanged. They form a subgroup and we study its influence upon $\Xi_2, \cdots, \Xi_{n-1}$. The next step would consist in singling out Ξ_2. By induction we thus obtain a finite number of subsidiary linear inequalities each concerned with the four components of one of the variables $\gamma_{12}, \cdots, \gamma_{n-1,n}$ only, and by them we define the fundamental domain G_{A}' in G_{A} for the group $\{J_{\mathrm{A}}\}$.

I set little store by this whittling down of G_{A} to G_{A}'. It seems less artificial to operate with the whole cells Z; in doing so one has to keep in mind that the modular group in its influence upon Z matters only modulo $\{J_Z\}$.

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

THE GENERAL INVARIANT THEORY OF IRREGULAR ANALYTIC ARCS OR ELEMENTS

BY

EDWARD KASNER AND JOHN DeCICCO

Introduction. In this paper, we shall begin the study of the invariant theory of the most general irregular analytic arc in the geometry based on the infinite group G of arbitrary regular point transformations. Our results are valid for the group of real point transformations of the real plane; or for the group of complex point transformations of the complex plane. Kasner has developed the corresponding theory in the conformal geometry of the complex plane(1). The present paper opens up a new aspect of restricted topology.

Our subject is then the equivalence theory of a *single* arc or curve. When can one analytic arc be converted into another analytic arc by an arbitrary regular point transformation of the plane? It is apparently implied, in the current literature, that there is no problem here. For any curve (it is implied) can be converted into any other, in particular, into the x-axis. But this is based on the assumption that the arcs are regular. If we give up this assumption, we have actual problems which certainly seem worthy of treatment. Our subject is therefore *the invariant theory of a general irregular analytic arc under the group G of arbitrary regular point transformations.*

More exactly, the configuration we shall discuss is not simply an analytic arc but rather that arc together with a specific point of the arc. This compound configuration we shall term an *analytic element*. It consists of a point—the base point which shall be taken as the origin throughout this paper—and an analytic arc through the point. It may be described also as a *differential element of infinite order*(2).

The most general analytic element, if the given point o is taken as origin, is represented by writing x and y as integral power series in a parameter t without the constant terms. If the parameter t is eliminated, then y is found as a series in x which may proceed according to *integral* or *fractional* powers

Presented to the Society, January 1, 1941, under the title *The classification of analytic arcs or elements under the group of arbitrary point transformations*; received by the editors January 2, 1941.

(1) Kasner, *Conformal classification of analytic arcs or elements; Poincaré's local problem of conformal geometry*, these Transactions, vol. 16 (1915), pp. 333–349.

(2) Kasner has introduced elsewhere the concept of divergent differential element of infinite order. This corresponds to a divergent power series and may be represented by a nonanalytic arc having specified values for all the successive derivatives. Thus to every power series corresponds a geometric entity which may be real or imaginary, regular or irregular, convergent or divergent. This entity is the most general differential element. If it is convergent, we call it an analytic element (regular or irregular), or more loosely, an analytic arc or a curve.

of x. If fractional exponents enter and cannot be avoided by interchanging x and y (this will then necessarily be the case for *any* choice of rectangular axes), we shall call the element *irregular*(3); otherwise the element is *regular*.

Our new problem is to classify with respect to the group G of arbitrary regular point transformations all irregular analytic elements.

It is obvious that *all* regular elements are equivalent under the group G of arbitrary point transformations. That is, all complex regular elements (both real and imaginary) are equivalent under the group G of complex point transformations; whereas all real regular elements are equivalent under the group G of all real point transformations. Any regular element may be reduced to the canonical form $y=0$ (the x-axis together with the origin as base point).

But for irregular elements, the results are quite complicated. (See the table on page 235.) It is clear, for example, that the cuspidal element $y=x^{3/2}$ cannot be converted into the regular element $y=0$, nor into the irregular element $y=x^{4/3}$. For these curves differ qualitatively in the nature of the singular point at the origin. However, suppose the two proposed elements have the same kind of irregularity (in a sense to be later defined, depending on agreement of certain exponents—certain arithmetic invariants); will they necessarily be equivalent? If not, certain combinations of the coefficients will be invariant, that is, there will be absolute differential invariants. For example, we find that every differential element of the form

$$y = x^{3/2} + c_4x^{4/2} + c_5x^{5/2} + \cdots,$$

can be (formally) reduced to $y=x^{3/2}$. On the other hand, not every element of the form

$$y = x^{9/4} + c_{10}x^{10/4} + c_{11}x^{11/4} + \cdots,$$

can be reduced to $y=x^{9/4}$. Hence in the first type there are no invariants; in the second type there exists an invariant.

In general, irregular elements have absolute differential invariants; certain exceptions arise, namely, those in which the corresponding series in x proceeds according to powers of the square root, and the cube root, and three particular types of series in x which proceed according to powers of the fourth root. *All the other cases possess absolute differential invariants.*

Statement of results. We shall throughout this paper assume that our group G of arbitrary regular point transformations leaves invariant the fixed point o (the origin) of our analytic element and that it is regular in the neighborhood of o. Thus our group G is

(3) This is related to the concept of *cycle* used in the theory of algebraic curves, and to the more general theory of algebroid arcs. Topologic invariants of algebroid arcs are studied in papers by Brauner, Kähler, and Zariski; for references see the latter's paper, American Journal of Mathematics, vol. 54 (1932), pp. 453–465.

$$
(1)\qquad G:\quad
\begin{aligned}
X &= (\alpha_{01}x + \alpha_{02}x^2 + \cdots) + y(\alpha_{10} + \alpha_{11}x + \cdots)\\
&\quad + y^2(\alpha_{20} + \alpha_{21}x + \cdots) + \cdots,\\
Y &= (\beta_{01}x + \beta_{02}x^2 + \cdots) + y(\beta_{10} + \beta_{11}x + \cdots)\\
&\quad + y^2(\beta_{20} + \beta_{21}x + \cdots) + \cdots,
\end{aligned}
$$

where the jacobian $J=\alpha_{01}\beta_{10}-\alpha_{10}\beta_{01}\neq 0$. We therefore shall study the invariant theory of an analytic element (with the origin as base point) with respect to the group G as given by these equations.

Any analytic element with origin as base point may be defined by setting x and y equal to two power series in a parameter t without constant terms. Let $p\geqq 1$ be the minimum of the two exponents of the two leading terms of the two power series in t. By interchanging the coordinates x and y appropriately, we can always arrange that the exponent of the leading term of the power series defining the abscissa x shall be our number p. Therefore any analytic element may be written in the form

$$
(2)\qquad y = c_p x^{p/p} + c_{2p}x^{2p/p} + \cdots + c_{rp}x^{rp/p} + c_q x^{q/p} + c_{q+1}x^{(q+1)/p} + \cdots,
$$

where q is the first qth power of the pth root of x which is *not* a multiple of p. That is, the integer $q=rp+s$ is such that the integers r and s satisfy the inequalities $r\geqq 1$ and $0<s<p$.

If $c_q\neq 0$, our element is said to be *irregular*. Otherwise our element is said to be *regular*. Thus an element is regular if $p=1$ (and hence no such term $c_q x^{q/p}$ can appear in our power series); or if $p\geqq 2$ and if $c_q=0$ (the coefficient of any actual fractional power of x is zero).

For an irregular element the integer $p\geqq 2$ is called the *index* and the integer $q=rp+s$, $r\geqq 1$, $0<s<p$, is termed the *rank*. All irregular elements obtained by taking arbitrary values of the coefficients, but fixing the values of both p and q, we shall define as forming the *single species* (p, q). It is proved that the index p and the rank q are *arithmetic invariants*. Our main result follows:

Absolute differential invariants, that is, functions of the coefficients unaltered by the group G of arbitrary point transformations, exist for all irregular species (p, q) *except in the cases* (4, 5), (4, 6), (4, 7), (3, q), *and* (2, q). *The species* (4, 5) *may be divided into two distinct sets, and the species* (4, 7) *may be separated into three distinct sets; the elements of any one of these sets are all equivalent to each other. The species* (4, 6) *and* (3, q) *may be separated into a denumerably infinite number of such distinct sets. All the elements of the species* (2, q) *are equivalent to each other—the canonical form is* $y=x^{q/2}$. *Finally all the regular elements are obviously transformable into each other, the standard form being* $y=0$.

The species $(p, q=rp+s)$ for which $r\geqq 2$ and $0<s<p-2$, or for which $r=1$ and $2<s<p-2$, possesses an absolute differential invariant of order

$q+2$ (and none of lower order). This invariant involves only the coefficients c_q, c_{q+1} and c_{q+2}.

The species $(p, q=rp+s)$ for which $r\geqq 2$ and $1<s=p-2$, or for which $r=1$ and $3<s=p-2$, possesses an absolute differential invariant of order $q+3$. This involves only the coefficients c_q, c_{q+1}, and c_{q+3}.

The species $(p, q=rp+s)$ for which $r\geqq 2$ and $2<s=p-1$, or for which $r=1$ and $3<s=p-1$, possesses an absolute differential invariant of order $q+3$. This involves only the coefficients c_q, c_{q+2}, and c_{q+3}.

The species $(p, q=p+1)$ for which $p>4$ possesses an absolute differential invariant of order $q+3=p+4$. This contains only the coefficients c_{p+1}, c_{p+2}, c_{p+3}, and c_{p+4}.

The species $(p, q=p+2)$ for which $p>5$ possesses an absolute differential invariant of order $q+3=p+5$. This contains only the coefficients c_{p+2}, c_{p+3}, c_{p+4}, and c_{p+5}.

The species (5, 7) possesses an absolute differential invariant of order 11. This contains only the coefficients c_7, c_8, c_9, and c_{11}.

Finally the species (5, 8) possesses an absolute differential invariant of order 12. This involves only the coefficients c_8, c_9, c_{11}, and c_{12}.

The following table exhibits some of the results in detail.

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2	*		*		*		*		*		*		*		*	
3		*	*		†	†		†	†		†	†		†	†	
4			†	‡	†		11	13	14		15	17	18		19	21
5				9	11	12	12		13	14	16	17		18	19	21
6					10	11	11	13	14		15	16	17	19	20	
7						11	12	12	13	15	16		17	18	19	20
8							12	13	13	14	15	17	18		19	20
9								13	14	14	15	16	17	19	20	
10									14	15	15	16	17	18	19	21

Here the species is determined by the value of p in the left column and the value of q in the top row. The blank spaces denote the fact that there are no such species. In the body of the table we find the order of the absolute differential invariants discussed above. The asterisk indicates that the corresponding species has no absolute invariant and further that all members of that species are equivalent under the group G of arbitrary point transformations. The dagger indicates that there are no absolute differential invariants but that the members are not all equivalent: there exists a certain relative differential invariant, hence there is a division of such a species into a finite number of subspecies, all distinct with respect to the group G of arbitrary

point transformations. On the other hand, the double dagger indicates that there exists a certain arithmetic invariant and hence the species (4, 6) can be divided into an infinite number of subspecies.

1. **Discussion of the regular elements.** Every regular element: $y=c_1x+c_2x^2+\cdots$, is carried by the point transformation: $X=x$, $Y=y-(c_1x+c_2x^2+\cdots)$, into the x-axis: $y=0$. Hence

THEOREM 1. *All regular analytic elements are equivalent under the group G of arbitrary point transformations. The normal form of a regular analytic element is $y=0$ (the x-axis together with the origin as base point).*

Our purpose in stating this elementary result is to give a complete classification of all elements both regular and irregular. Henceforth we only have to confine our attention to irregular elements. In the next section, we shall discuss the index p and the rank q of any irregular element.

2. **The invariance of the index p and the rank q.** First let us note that by the point transformation

$$X = x, \qquad Y = y - (c_px + c_{2p}x^2 + c_{3p}x^3 + \cdots + c_{rp}x^r), \tag{3}$$

our irregular element (2) is carried into the irregular analytic element

$$y = c_qx^{q/p} + c_{q+1}x^{(q+1)/p} + \cdots, \qquad c_q \neq 0, \tag{4}$$

where the coefficients $c_q, c_{q+1}, c_{q+2}, \cdots$ of this element are identical with the corresponding ones of our irregular element (2). Thus the transformation (3) eliminates the first r integral powers of x.

Applying any transformation of our group G to this element (4), we find as the parametric equations of our new element

$$\begin{aligned} X &= \alpha_{01}t^p + \alpha_{02}t^{2p} + \cdots + \alpha_{0r}t^{rp} + \alpha_{10}c_qt^q + \cdots, \\ Y &= \beta_{01}t^p + \beta_{02}t^{2p} + \cdots + \beta_{0r}t^{rp} + \beta_{10}c_qt^q + \cdots. \end{aligned} \tag{5}$$

Since the jacobian of the transformation is not zero, at least one of the quantities α_{01} or β_{01} is not zero. Hence the number p is the minimum exponent in the two power series defining our new element and therefore must be its index. Thus *the group G preserves the index p.*

By interchanging the coordinates X and Y appropriately, we can always arrange so that $\alpha_{01}\neq 0$. Upon setting $X=T^p$, the first of the preceding equations defines t as an integral power series in T. This power series must be of the form

$$\begin{aligned} t = A_1T &+ A_{p+1}T^{p+1} + A_{2p+1}T^{2p+1} + \cdots \\ &+ A_{rp-p+1}T^{rp-p+1} + A_{q-p+1}T^{q-p+1} + \cdots, \end{aligned} \tag{6}$$

where the exponents after the last written term increase by one (as far as we know).

Substituting this into the first of equations (5), we obtain

$$
(7)\qquad
\begin{aligned}
\alpha_{01}A_1^{p} &= 1,\\
p\alpha_{01}A_1^{p-1}A_{p+1} + \alpha_{02}A_1^{2p} &= 0,\\
p\alpha_{01}A_1^{p-1}A_{2p+1} + \cdots + \alpha_{03}A_1^{3p} &= 0,\\
\cdots\cdots\cdots\cdots\cdots\cdots &,\\
p\alpha_{01}A_1^{p-1}A_{rp-p+1} + \cdots + \alpha_{0r}A_1^{rp} &= 0,\\
p\alpha_{01}A_1^{p-1}A_{q-p+1} + \alpha_{10}c_qA_1^{q} &= 0,\\
\cdots\cdots\cdots\cdots\cdots\cdots &.
\end{aligned}
$$

This system of equations determines the A's in terms of the α's. We note that $A_1 \neq 0$. It is also important to note that the last written equation contains only *two* terms.

If $(\alpha_{01})^{1/p}$ denotes one and only one pth root of α_{01}, then it is easily seen that these equations determine a one-to-one correspondence between $(\alpha_{01}, \alpha_{02}, \cdots, \alpha_{0r})$ and $(A_1, A_{p+1}, A_{2p+1}, \cdots, A_{rp-p+1})$. Hence we may replace the preceding set of unknown α's by the latter set of unknown A's.

Also let us note the following result

$$
(8)\qquad J = \alpha_{01}\beta_{10} - \alpha_{10}\beta_{01} = \frac{1}{c_qA_1^{q+1}}\left(\beta_{10}A_1^{q-p+1}c_q + p\beta_{01}A_{q-p+1}\right) \neq 0.
$$

This is needed for the completion of our argument.

Now substituting (6) into the second of equations (5), our new element (5) may be written in the form

$$
(9)\qquad
\begin{aligned}
Y = \beta_{01}A_1^{p}X^{p/p} &+ (\beta_{02}A_1^{2p} + p\beta_{01}A_1^{p-1}A_{2p+1})X^{2p/p} + (\beta_{03}A_1^{3p} + \cdots)X^{3p/p}\\
&+ \cdots + (\beta_{0r}A_1^{rp} + \cdots)X^{rp/p}\\
&+ A_1^{p-1}[\beta_{10}A_1^{q-p+1}c_q + p\beta_{01}A_{q-p+1}]X^{q/p} + \cdots.
\end{aligned}
$$

This equation shows that the first power of the pth root of X which is not a multiple of p is the integer q. The coefficient of this power is not zero because of (8). Hence *the rank q is also preserved.*

THEOREM 2. *The index p and the rank q of an irregular element are both arithmetic invariants under the group G of arbitrary point transformations.*

This theorem justifies our definition of the species (p, q). That is, under the group of arbitrary point transformations any species (p, q) of irregular elements is carried into itself.

By the above equation, it is found that the subgroup G' of the group G which carries the element (4) into one of the same form is

$$(10)\quad G':\quad \begin{aligned} X &= (\alpha_{01}x + \alpha_{02}x^2 + \cdots) + y(\alpha_{10} + \alpha_{11}x + \cdots) + \cdots, \\ Y &= (\beta_{0,r+1}x^{r+1} + \beta_{0,r+2}x^{r+2} + \cdots) + y(\beta_{10} + \beta_{11}x + \cdots) + \cdots, \end{aligned}$$

where

$$\alpha_{01}\beta_{10} \neq 0.$$

Since our group G may be factored into the product of the group (3) by the group G', and since the group (3) preserves the coefficients $c_q, c_{q+1}, \cdots$, it is necessary merely to study the invariant theory of the group G'.

In the following, we shall find the different invariants of lowest order of our irregular elements. For this purpose, we shall find it convenient to classify our elements according to the types of invariants which arise. See §§3 to 14 which follow. Some of the twelve classes are simple and some are complicated. The invariants found vary greatly in structure it will be observed.

3. **The discussion of the species (p, $q=rp+s$) for which $r\geqq 2$ and $0<s<p-2$, or for which $r=1$ and $2<s<p-2$.** Substituting (6) into the second of equations (10), we find because of our inequalities the following transformation formulas between (c_q, c_{q+1}, c_{q+2}) and (C_q, C_{q+1}, C_{q+2}):

$$(11)\qquad C_q = \beta_{10}A_1^q c_q, \qquad C_{q+1} = \beta_{10}A_1^{q+1}c_{q+1}, \qquad C_{q+2} = \beta_{10}A_1^{q+2}c_{q+2}.$$

These equations immediately yield the following result.

THEOREM 3. *The species (p, $q=rp+s$) for which $r\geqq 2$ and $0<s<p-2$, or for which $r=1$ and $2<s<p-2$ possesses the absolute differential invariant of lowest order*

$$(12)\qquad \frac{c_q c_{q+2}}{c_{q+1}^2}.$$

The order of our invariant is $q+2$, its weight is $2q+2$, and its degree is 2.

4. **The discussion of the species (p, $q=rp+s$) for which $r\geqq 2$ and $1<s=p-2$, or for which $r=1$ and $3<s=p-2$.** Again because of our inequalities, we obtain the following transformation formulas between (c_q, c_{q+1}, c_{q+3}) and (C_q, C_{q+1}, C_{q+3}):

$$(13)\qquad C_q = \beta_{10}A_1^q c_q, \qquad C_{q+1} = \beta_{10}A_1^{q+1}c_{q+1}, \qquad C_{q+3} = \beta_{10}A_1^{q+3}c_{q+3}.$$

Therefore

THEOREM 4. *This species possesses the absolute invariant*

$$(14)\qquad \frac{c_q^2 c_{q+3}}{c_{q+1}^3}.$$

The order is $q+3$, weight is $3q+3$, and degree is 3.

5. **The discussion of the species (p, $q=rp+s$) for which $r\geqq 2$ and $2<s=p-1$, or for which $r=1$ and $3<s=p-1$.** Because of these inequalities, the following transformation formulas between (c_q, c_{q+2}, c_{q+3}) and (C_q, C_{q+2}, C_{q+3}) are obtained:

$$C_q = \beta_{10}A_1^q c_q, \quad C_{q+2} = \beta_{10}A_1^{q+2} c_{q+2}, \quad C_{q+3} = \beta_{10}A_1^{q+3} c_{q+3}. \tag{15}$$

THEOREM 5. *This species possesses the absolute invariant*

$$\frac{c_q^2 c_{q+3}}{c_{q+2}^3}. \tag{16}$$

The order is $q+3$, weight is $3q+6$, and degree is 3.

6. **Discussion of the species (p, $q=p+1$) for which $p>4$.** For the case $q=p+1$, we find that the parametric equations (5) for our transformed element assume the form

$$\begin{aligned} X &= \alpha_{01}t^p + \alpha_{10}(c_{p+1}t^{p+1} + c_{p+2}t^{p+2} + c_{p+3}t^{p+3} + \cdots) + \cdots, \\ Y &= \beta_{10}(c_{p+1}t^{p+1} + c_{p+2}t^{p+2} + c_{p+3}t^{p+3} + c_{p+4}t^{p+4} + \cdots) + \cdots. \end{aligned} \tag{17}$$

Let $X=T^p$. The first of the preceding equations defines t as an integral power series in T. This power series must be of the form

$$t = A_1T + A_2T^2 + A_3T^3 + \cdots, \qquad A_1 \neq 0. \tag{18}$$

Substituting this value of t into the first of the preceding equations, we obtain the following system of equations:

$$\begin{aligned} &\alpha_{01}A_1^p = 1, \qquad p\alpha_{01}A_1^{p-1}A_2 + \alpha_{10}A_1^{p+1}c_{p+1} = 0, \\ &\alpha_{01}\left[pA_1^{p-1}A_3 + \frac{p}{2}(p-1)A_1^{p-2}A_2^2\right] \\ &\qquad + \alpha_{10}[(p+1)A_1^pA_2c_{p+1} + A_1^{p+2}c_{p+2}] = 0, \\ &\alpha_{01}\left[pA_1^{p-1}A_4 + p(p-1)A_1^{p-2}A_2A_3 + \frac{p}{6}(p-1)(p-2)A_1^{p-3}A_2^3\right] \\ &\qquad + \alpha_{10}\left[\left\{(p+1)A_1^pA_3 + \frac{p}{2}(p+1)A_1^{p-1}A_2^2\right\}c_{p+1} \right. \\ &\qquad\qquad \left. + (p+2)A_1^{p+1}A_2c_{p+2} + A_1^{p+3}c_{p+3}\right] = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{19}$$

Now if $c_{p+1}\neq 0$, we find that A_1 and A_2 are arbitrary, but A_3 and A_4 are dependent, being given by

$$
(20)\qquad
\begin{aligned}
A_3 &= \frac{1}{2}(p+3)\frac{A_2^2}{A_1} + A_1A_2\frac{c_{p+2}}{c_{p+1}},\\
A_4 &= \frac{1}{3}(p+2)(p+4)\frac{A_2^3}{A_1^2} + (p+4)A_2^2\frac{c_{p+2}}{c_{p+1}} + A_1^2A_2\frac{c_{p+3}}{c_{p+1}}.
\end{aligned}
$$

Substituting (18) into the second of equations (17), we discover the following transformation formulas between $(c_{p+1}, c_{p+2}, c_{p+3}, c_{p+4})$ and $(C_{p+1}, C_{p+2}, C_{p+3}, C_{p+4})$:

$$
\begin{aligned}
C_{p+1} &= \beta_{10}A_1^{p+1}c_{p+1}, \qquad C_{p+2} = \beta_{10}\{A_1^{p+2}c_{p+2} + (p+1)A_1^pA_2c_{p+1}\},\\
C_{p+3} &= \beta_{10}\Big\{A_1^{p+3}c_{p+3} + (p+2)A_1^{p+1}A_2c_{p+2}\\
&\qquad + \Big[(p+1)A_1^pA_3 + \frac{p}{2}(p+1)A_1^{p-1}A_2^2\Big]c_{p+1}\Big\},
\end{aligned}
$$

$$
(21)\qquad
\begin{aligned}
C_{p+4} &= \beta_{10}\Big\{A_1^{p+4}c_{p+4} + (p+3)A_1^{p+2}A_2c_{p+3}\\
&\qquad + \Big[(p+2)A_1^{p+2}A_3 + \frac{1}{2}(p+1)(p+2)A_1^pA_2^2\Big]c_{p+2}\\
&\qquad + \Big[(p+1)A_1^pA_4 + p(p+1)A_1^{p-1}A_2A_3\\
&\qquad\qquad + \frac{p}{6}(p-1)(p+1)A_1^{p-2}A_2^3\Big]c_{p+1}\Big\}.
\end{aligned}
$$

Eliminating β_{10}, A_1, A_2, A_3, and A_4 from these and the preceding equations, we find

THEOREM 6. *The species $(p, q=p+1)$ for which $p>4$ possesses the absolute invariant*

$$
(22)\qquad \frac{[(p+1)^2c_{p+1}^2c_{p+4} - 2(p+1)(p+2)c_{p+1}c_{p+2}c_{p+3} + \frac{1}{3}(p+2)(3p+5)c_{p+2}^3]^2}{[(p+1)c_{p+1}c_{p+3} - \frac{1}{2}(2p+3)c_{p+2}^2]^3}.
$$

The order is $q+3=p+4$; weight is $3q+3=3p+6$, and degree is 6.

7. **Discussion of the species $(p, q=p+2)$ for which $p>5$.** The calculations for our transformed element will be given by the equations of the preceding section if we assume $c_{p+1}=0$ but $c_{p+2}\neq 0$.

Under this restriction we find that $A_1\neq 0$ and A_3 are arbitrary but A_2 and A_4 are dependent, being given by

$$A_2 = 0, \qquad A_4 = A_1A_3\frac{c_{p+3}}{c_{p+2}}\cdot \tag{23}$$

The transformation formulas between $(c_{p+2}, c_{p+3}, c_{p+4}, c_{p+5})$ and $(C_{p+2}, C_{p+3}, C_{p+4}, C_{p+5})$ are

$$\begin{aligned} C_{p+2} &= \beta_{10}A_1^{p+2}c_{p+2}, \qquad C_{p+3} = \beta_{10}A_1^{p+3}c_{p+3}, \\ C_{p+4} &= \beta_{10}[A_1^{p+4}c_{p+4} + (p+2)A_1^{p+3}A_3c_{p+2}], \\ C_{p+5} &= \beta_{10}[A_1^{p+5}c_{p+5} + (p+3)A_1^{p+2}A_3c_{p+3} + (p+2)A_1^{p+1}A_4c_{p+2}]. \end{aligned} \tag{24}$$

Performing the elimination of β_{10}, A_1, A_3, and A_4 from these equations, we find

THEOREM 7. *The species* $(p, q=p+2)$ *for which* $p>5$ *possesses the absolute invariant*

$$\frac{c_{p+2}[(p+2)c_{p+2}c_{p+5} - (2p+5)c_{p+3}c_{p+4}]}{c_{p+3}^3}\cdot \tag{25}$$

The order is $q+3=p+5$, weight is $3q+3=3p+9$, and degree is 3.

8. **Discussion of the species** (5, 7). For the species $(p=5, q=7)$, it is found that the parametric equations (5) of the transformed element assume the form

$$\begin{aligned} X &= \alpha_{01}t^5 + \alpha_{10}(c_7t^7 + c_8t^8 + c_9t^9 + \cdots) + \cdots, \\ Y &= \beta_{10}(c_7t^7 + c_8t^8 + c_9t^9) + (\beta_{02} + \beta_{10}c_{10})t^{10} + \beta_{10}c_{11}t^{11} + \cdots. \end{aligned} \tag{26}$$

Let $X=T^5$. The first of the preceding equations defines t as an integral power series in T. This power series must be of the form

$$t = A_1T + A_3T^3 + A_4T^4 + A_5T^5 + \cdots, \qquad A_1 \neq 0. \tag{27}$$

Replacing t by this in the first of equations (26), we find

$$\begin{aligned} \alpha_{01}A_1^5 = 1, \qquad & 5\alpha_{01}A_1^4A_3 + \alpha_{10}A_1^7c_7 = 0, \\ & 5\alpha_{01}A_1^4A_4 + \alpha_{10}A_1^8c_8 = 0, \\ \alpha_{01}(5A_1^4A_5 + 10A_1^3A_3^2) &+ \alpha_{10}(7A_1^6A_3c_7 + A_1^9c_9) = 0, \\ \cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{28}$$

If $c_7 \neq 0$, these equations show that A_1 and A_3 are arbitrary, but A_4 and A_5 are dependent being given by

$$A_4 = A_1A_3\frac{c_8}{c_7}, \qquad A_5 = \frac{5A_3^2}{A_1} + A_1^2A_3\frac{c_9}{c_7}\cdot \tag{29}$$

Substituting (27) into the second of equations (26), we obtain the follow-

ing transformation formulas between (c_7, c_8, c_9, c_{11}) and (C_7, C_8, C_9, C_{11}):

$$
\begin{aligned}
&C_7 = \beta_{10} A_1^7 c_7, \qquad C_8 = \beta_{10} A_1^8 c_8,\\
(30) \qquad &C_9 = \beta_{10}[A_1^9 c_9 + 7A_1^6 A_3 c_7],\\
&C_{11} = \beta_{10}[c_{11} A_1^{11} + 9A_1^8 A_3 c_9 + 8A_1^7 A_4 c_8 + (7A_1^6 A_5 + 21 A_1^5 A_3^2) c_7].
\end{aligned}
$$

THEOREM 8. *The species* (5, 7) *possesses the absolute invariant*

$$
(31) \qquad \frac{c_7(c_7^2 c_{11} - 8c_7 c_9^2 - 8c_8^2 c_9)}{c_8^4}.
$$

The order is 11, weight is 32, and degree is 4.

9. **Discussion of the spaces** (5, 8). The material for our transformed element will be given by the equations of the preceding section if we assume $c_7 = 0$ but $c_8 \neq 0$.

Since $c_8 \neq 0$, we find by equations (28) that A_1 and A_4 are arbitrary but A_3 and A_5 are dependent being given by

$$
(32) \qquad A_3 = 0, \qquad A_5 = A_1 A_4 \frac{c_9}{c_8}.
$$

By (30), we see that for this case the transformation formulas between $(c_8, c_9, c_{11}, c_{12})$ and $(C_8, C_9, C_{11}, C_{12})$ are

$$
\begin{aligned}
&C_8 = \beta_{10} A_1^8 c_8, \qquad C_9 = \beta_{10} A_1^9 c_9,\\
(33) \qquad &C_{11} = \beta_{10}[A_1^{11} c_{11} + 8A_1^7 A_4 c_8],\\
&C_{12} = \beta_{10}[A_1^{12} c_{12} + 9A_1^8 A_4 c_9 + 8A_1^7 A_5 c_8].
\end{aligned}
$$

THEOREM 9. *The species* (5, 8) *possesses the absolute invariant*

$$
(34) \qquad \frac{c_8^2(8c_8 c_{12} - 9c_9 c_{11})}{c_9^4}.
$$

The order is 12, weight is 36, and degree is 4.

10. **Discussion of the species** (4, 5). In the following, we shall show that this species only possesses a *relative* differential invariant. This species therefore can be separated into two distinct sets, the elements of any one set being equivalent.

Part of the work for finding the coefficients of the transformed element is the same as that performed for the species $(p, q = p+1)$ for $p > 4$ (§6). As a matter of fact the first three of equations (21) are valid. From these it follows that *the species* (4, 5) *possesses* $(10c_5c_7 - 11c_6^2)^2/c_5^5$ *as a relative differential invariant.*

In the following, we shall prove that there are no more relative differential invariants. This is done by reducing our species (4, 5) to the canonical form.

First let us note that the transformation

$$X = x + \frac{4c_6}{5c_5^2}\,y, \qquad Y = y \tag{35}$$

will convert our element (4) of species (4, 5) into an element whose equation is of the form

$$y = d_5x^{5/4} + d_7x^{7/4} + d_8x^{8/4} + \cdots, \tag{36}$$

where the d's are certain functions of the c's. Our relative differential invariant for this element is $100d_7^2/d_5^3$ (since $d_5 \neq 0$). Thus we have to discuss this element according as $d_7 \neq 0$ or $d_7 = 0$.

First let us consider our element (36) for which $d_7 \neq 0$. For this case, we shall demonstrate the (formal) existence of a transformation which is the *inverse* of a transformation of the form

$$\begin{aligned} X &= \alpha_{01}x + \alpha_{02}x^2 + \alpha_{03}x^3 + \cdots, \\ Y &= (\beta_{02}x^2 + \beta_{03}x^3 + \beta_{04}x^4 + \cdots) + y(\beta_{10} + \beta_{11}x + \beta_{12}x^2 + \beta_{13}x^3 + \cdots) \\ &\quad + y^2(\beta_{20} + \beta_{21}x + \beta_{22}x^2 + \beta_{23}x^3 + \cdots), \end{aligned} \tag{37}$$

which will carry our element into the canonical form $y = x^{5/4} + x^{7/4}$. Therefore it must be shown that the α_{0j} and the β_{ij} ($i = 0, 1, 2$) can be determined so that our transformation carries the canonical form into the element (36).

Now if we perform this transformation on the canonical form $y = x^{5/4} + x^{7/4}$, the parametric form of our transformed element is

$$\begin{aligned} X &= \alpha_{01}t^4 + \alpha_{02}t^8 + \alpha_{03}t^{12} + \cdots, \\ Y &= [\beta_{02}t^8 + (\beta_{03} + 2\beta_{20})t^{12} + (\beta_{04} + 2\beta_{21})t^{16} + \cdots] \\ &\quad + [\beta_{20}t^{10} + (\beta_{21} + \beta_{20})t^{14} + (\beta_{22} + \beta_{21})t^{18} + \cdots] \\ &\quad + [\beta_{10}(t^5 + t^7) + \beta_{11}(t^9 + t^{11}) + \beta_{12}(t^{13} + t^{15}) + \cdots]. \end{aligned} \tag{38}$$

The first of these equations defines t as an integral power series in T if we let $X = T^4$. This power series must be of the form $t = A_1T + A_5T^5 + A_9T^9 + \cdots$ where $A_1 \neq 0$. If we let $A_1^{1/4}$ denote a definite fourth root of A_1, it may be established that there exists a one-to-one correspondence between $(\alpha_{01}, \alpha_{02}, \cdots, \alpha_{0n}, \cdots)$ and $(A_1, A_5, \cdots, A_{4n-3}, \cdots)$. Therefore we may replace the unknown α's by the unknown A's.

In the second of the preceding equations, let us replace the coefficient of t^{4n+4} by B_{4n+4}, the coefficient of t^{4n+6} by B_{4n+6}, and the coefficient of t^{4n+1} by B_{4n+1} where $n = 1, 2, 3, \cdots$. Then obviously there exists a one-to-one correspondence between the β_{ij} and the B_k. Thus we may substitute the unknown B's for the unknown β's.

By the preceding remarks, our problem then is to determine the A's and the B's so that the equations

$$
\begin{aligned}
&X = T^4, \qquad t = A_1T + A_5T^5 + A_9T^9 + \cdots, \\
(39)\quad &Y = [B_8t^8 + B_{12}t^{12} + B_{16}t^{16} + \cdots] + [B_{10}t^{10} + B_{14}t^{14} + B_{18}t^{18} + \cdots] \\
&\qquad + [B_5(t^5 + t^7) + B_9(t^9 + t^{11}) + B_{13}(t^{13} + t^{15}) + \cdots],
\end{aligned}
$$

shall represent the element (36) upon eliminating the parameters t and T.

Performing this elimination, we find our equivalence equations to be

$$
(40.1)\quad
\begin{gathered}
B_8A_1^8 = d_8, \qquad B_{12}A_1^{12} + 8B_8A_1^7A_5 = d_{12}, \\
\cdots\cdots\cdots\cdots\cdots\cdots, \\
B_{4n+4}A_1^{4n+4} + (\text{lower } B_{4k}) = d_{4n+4}, \\
\cdots\cdots\cdots\cdots\cdots\cdots;
\end{gathered}
$$

$$
(40.2)\quad
\begin{gathered}
B_{10}A_1^{10} = d_{10}, \qquad B_{14}A_1^{14} + 10B_{10}A_1^9A_5 = d_{14}, \\
\cdots\cdots\cdots\cdots\cdots\cdots, \\
B_{4n+6}A_1^{4n+6} + (\text{lower } B_{4k+6}) = d_{4n+6}, \\
\cdots\cdots\cdots\cdots\cdots\cdots;
\end{gathered}
$$

$$
(40.3)\quad
\begin{gathered}
B_5A_1^5 = d_5, \qquad 5A_1^4A_5B_5 + B_9A_1^9 = d_9, \\
B_5A_1^7 = d_7, \qquad 7A_1^6A_5B_5 + B_9A_1^{11} = d_{11}, \\
\cdots\cdots\cdots\cdots\cdots\cdots, \\
5A_1^4A_{4n-3}B_5 + B_{4n+1}A_1^{4n+1} + (\text{lower } A_{4k-3} \text{ and } B_{4k+1}) = d_{4n-3}, \\
7A_1^6A_{4n-3}B_5 + B_{4n+1}A_1^{4n+3} + (\text{lower } A_{4k-3} \text{ and } B_{4k+1}) = d_{4n-1}, \\
\cdots\cdots\cdots\cdots\cdots\cdots.
\end{gathered}
$$

The first pair of the equations (40.3) shows that $A_1 \neq 0$ and $B_5 \neq 0$ are uniquely determined. The second pair demonstrates that A_5 and B_9 are uniquely determined since the determinant of the coefficients $-2A_1^{15}B_5 \neq 0$. The nth pair proves that A_{4n-3} and B_{4n+1} are uniquely evaluated since the determinant of the coefficients $-2A_1^{4n+7}B_5 \neq 0$. Thus by induction, all the A_{4n-3} and the B_{4n+1} are uniquely found.

The equations (40.1) and (40.2) uniquely evaluate the B_{4n+4} and the B_{4n+6}, respectively. Hence we are able to find unique values for all the A's and the B's and therefore (39) actually represents our element (36). Thus *the species* (4, 5) *for which the relative differential invariant* $(10c_5c_7 - 11c_6^2)^2/c_5^5 \neq 0$ *possesses no additional invariants. The canonical form is* $y = x^{5/4} + x^{7/4}$.

Finally let us consider the case for which our relative differential invariant is zero. Then for our element (36) we must have $d_7=0$. The *inverse* of the transformation

$$X = x, \qquad Y = y + d_8x^2 - \frac{d^9}{d_5}xy - \frac{d_{10}}{d_5^2}y^2, \tag{41}$$

will then carry our element (36) into one of the form

$$y = e_5x^{5/4} + e_{11}x^{11/4} + e_{12}x^{12/4} + \cdots, \tag{42}$$

where the e's are certain functions of the d's and hence it follows that they are functions of the c's.

Next the transformation which is the *inverse* of the correspondence

$$X = x + \frac{4e_{11}}{5e_5^3}y, \qquad Y = y, \tag{43}$$

will carry our element (42) into an element of the form

$$y = f_5x^{5/4} + f_{12}x^{12/4} + f_{13}x^{13/4} + \cdots, \tag{44}$$

where the f's are certain functions of the e's and hence of the c's.

Finally the *inverse* of the transformation

$$\begin{aligned} X = x, \qquad Y = &(f_{12}x^3 + f_{16}x^4 + f_{20}x^5 + \cdots) \\ &+ y(f_5 + f_{13}x^2 + f_{17}x^3 + f_{21}x^4 + \cdots) \\ &+ y^2(f_{14} + f_{18}x^2 + f_{22}x^3 + \cdots) \\ &+ y^3(f_{15} + f_{19}x + f_{23}x^2 + \cdots) \end{aligned} \tag{45}$$

will carry our element (44) into the element $y=x^{5/4}$. Thus *the species* (4, 5) *for which the relative differential invariant* $(10c_5c_7-11c_6^2)^2/c_5^2=0$ *possesses no additional invariants. The canonical form is* $y=x^{5/4}$.

Therefore we have proved the following result.

THEOREM 10. *The species* (4, 5) *possesses only the relative differential invariant*

$$\frac{(10c_5c_7 - 11c_6^2)^2}{c_5^5} \tag{46}$$

and no other differential or arithmetic invariants. Thus our species (4, 5) *may be separated into two distinct sets according as this relative differential invariant is not or is zero. The canonical forms of these two distinct sets are* $y=x^{5/4}+x^{7/4}$, *and* $y=x^{5/4}$, *respectively.*

11. **Discussion of the species** (4, 7). For this case, we see that the parametric equations (5) for our new element assume the form

$$
(47)\qquad
\begin{aligned}
X &= \alpha_{01}t^4 + \alpha_{10}c_7t^7 + (\alpha_{02} + \alpha_{10}c_8)t^8 + \alpha_{10}(c_9t^9 + c_{10}t^{10}) + \cdots,\\
Y &= \beta_{10}c_7t^7 + (\beta_{02} + \beta_{10}c_8)t^8 + \beta_{10}(c_9t^9 + c_{10}t^{10})\\
&\quad + (\beta_{11}c_7 + \beta_{10}c_{11})t^{11} + (\beta_{03} + \beta_{10}c_{12} + \beta_{11}c_8)t^{12}\\
&\quad + (\beta_{11}c_9 + \beta_{10}c_{13})t^{13} + \cdots.
\end{aligned}
$$

Let $X=T^4$. The first of these equations defines t as an integral power series in T. This power series must be of the form

$$
(48)\qquad t = A_1T + A_4T^4 + A_5T^5 + A_6T^6 + A_7T^7 + \cdots, \qquad A_1 \neq 0.
$$

Upon replacing t by this value in the first of equations (47), we obtain the following system of equations

$$
(49)\qquad
\begin{aligned}
\alpha_{01}A_1^4 = 1, \qquad 4\alpha_{01}A_1^3A_4 + \alpha_{10}A_1^7c_7 &= 0,\\
4\alpha_{01}A_1^3A_5 + (\alpha_{02} + \alpha_{10}c_8)A_1^8 &= 0,\\
4\alpha_{01}A_1^3A_6 + \alpha_{10}A_1^9c_9 &= 0,\\
\alpha_{01}(4A_1^3A_7 + 6A_1^2A_4^2) + \alpha_{10}(7A_1^6A_4c_7 + A_1^{10}c_{10}) &= 0,\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &.
\end{aligned}
$$

By these equations, we find, since $c_7\neq 0$, that A_1, A_4, and A_5 are arbitrary. But the values of A_6 and A_7 are dependent, being given by

$$
(50)\qquad
\begin{aligned}
A_6 &= \frac{c_9}{c_7}A_1^2A_4,\\
A_7 &= \frac{11A_4^2}{2A_1} + \frac{c_{10}}{c_7}A_1^3A_4.
\end{aligned}
$$

Upon substituting (48) into the second of equations (47), we obtain the following transformation formulas between $(c_7, c_8, c_9, c_{10}, c_{11}, c_{12}, c_{13})$ and $(C_7, C_8, C_9, C_{10}, C_{11}, C_{12}, C_{13})$:

$$
(51)\qquad
\begin{aligned}
C_7 &= \beta_{10}A_1^7c_7, \qquad C_8 = \beta_{10}A_1^8c_8 + \beta_{02}A_1^8,\\
C_9 &= \beta_{10}A_1^9c_9, \qquad C_{10} = \beta_{10}(A_1^{10}c_{10} + 7A_1^6A_4c_7),\\
C_{11} &= 8\beta_{02}A_1^7A_4 + \beta_{11}A_1^{11}c_7 + \beta_{10}(c_{11}A_1^{11} + 8A_1^7A_4c_8 + 7A_1^6A_5c_7),\\
C_{12} &= 8\beta_{02}A_1^7A_5 + \beta_{03}A_1^{12} + \beta_{11}A_1^{12}c_8\\
&\quad + \beta_{10}(A_1^{12}c_{12} + 9A_1^8A_4c_9 + 8A_1^7A_5c_8 + 7A_1^6A_6c_7),\\
C_{13} &= 8\beta_{02}A_1^7A_6 + \beta_{11}A_1^{13}c_9\\
&\quad + \beta_{10}[A_1^{13}c_{13} + 10A_1^9A_4c_{10} + 9A_1^8A_5c_9 + 8A_1^7A_6c_8\\
&\qquad\qquad + c_7(7A_1^6A_7 + 21A_1^5A_4^2)].
\end{aligned}
$$

By these equations, we discover that *the species* (4, 7) *possesses* c_9^3/c_7^4 *as a relative differential invariant. Moreover if this relative invariant is zero* ($c_9=0$), *then the species* (4, 7) *will possess* $(14c_7c_{13}-17c_{10}^2)/c_7^3$ *as a relative differential invariant.* In the remainder of this section, we shall demonstrate that there are no additional invariants. This is accomplished by reducing our element of species (4, 7) to the canonical form.

In the first place, it is observed that the transformation

$$X = x + \frac{4c_{10}}{7c_7^2}\,y, \qquad Y = y, \tag{52}$$

will carry our element (4) of species (4, 7) into an element of the form

$$y = d_7x^{7/4} + d_8x^{8/4} + d_9x^{9/4} + d_{11}x^{11/4} + \cdots, \tag{53}$$

where the d's are certain functions of the c's. Our first relative differential invariant for this element is d_9^3/d_7^4. Thus we have to discuss this element according as $d_9\neq 0$ or $d_9=0$.

Let us now consider the case where $d_9\neq 0$. By a very similar argument to the one given near the beginning of §10, it is very easy to demonstrate the (formal) existence of a unique transformation which is the *inverse* of a correspondence of the form (37) which will carry out element (53) into the canonical form $y=x^{7/4}+x^{9/4}$. Therefore *the species* (4, 7) *for which the relative differential invariant* $c_9^3/c_7^4\neq 0$ *possesses no additional invariants. The normal form is* $y=x^{7/4}+x^{9/4}$.

Next let us consider the case where $d_9=0$. For our element (53), our second relative differential invariant is $14d_{13}/d_7^2$. We next have to consider this element according as $d_{13}\neq 0$ or $d_{13}=0$.

We observe that in any case (whether $d_{13}\neq 0$ or $d_{13}=0$) that the transformation

$$X = x + \frac{4d_{11}}{7d_7}\,x^2, \qquad Y = y, \tag{54}$$

will carry our element (53) with $d_9=0$ into an element of the form

$$y = e_7x^{7/4} + e_8x^{8/4} + e_{12}x^{12/4} + e_{13}x^{13/4} + \cdots. \tag{55}$$

It remains to discuss this element according as $e_{13}\neq 0$ or $e_{13}=0$.

Let us now consider the case where $e_{13}\neq 0$. By an argument very similar to the one given near the beginning of §10, it is easy to prove the (formal) existence of a unique transformation which is the *inverse* of a correspondence of the form (37) which will carry our element (55) into the canonical form $y=x^{7/4}+x^{13/4}$. Therefore *the species* (4, 7) *for which* $c_9=0$ *and the relative differential invariant* $(14c_7c_{13}-17c_{10}^2)/c_7^3\neq 0$ *possesses no other invariants. The canonical form is* $y=x^{7/4}+x^{13/4}$.

Finally it remains to consider the case where $e_{13}=0$. It is observed that the transformation

$$X = x + \frac{4c_{17}}{7c_7^3}\,y, \qquad Y = y, \tag{56}$$

will carry our element (55) into one of the form

$$\begin{aligned} y = {} & f_7x^{7/4} + f_8x^{8/4} + f_{12}x^{12/4} + f_{14}x^{14/4} + f_{15}x^{15/4} + f_{16}x^{16/4} \\ & + f_{18}x^{18/4} + \cdots, \end{aligned} \tag{57}$$

where the f's are certain functions of the e's and therefore of the c's.

The transformation whose inverse is

$$\begin{aligned} X = x, \qquad Y = {} & (f_8x^2 + f_{12}x^3 + f_{16}x^4 + \cdots) \\ & + y(f_7 + f_{15}x^2 + f_{19}x^3 + \cdots) \\ & + y^2(f_{14} + f_{18}x + f_{22}x^2 + \cdots \\ & + y^3(f_{21} + f_{25}x + f_{29}x^2 + \cdots), \end{aligned} \tag{58}$$

will carry our element (57) into the normal form $y=x^{7/4}$. Therefore *the species* (4, 7) *for which* $c_9=0$ *and* $14c_7c_{13}-17c_{10}^2=0$ *possesses no additional invariants. The canonical form is* $y=x^{7/4}$.

We thus may make the following statement.

THEOREM 11. *The species* (4, 7) *possesses the relative differential invariant*

$$c_9^3/c_7^4. \tag{59}$$

If this relative differential invariant is zero, then that set of those elements of this species for which $c_9=0$ *possesses the additional relative differential invariant*

$$\frac{14c_7c_{13} - 17c_{10}^2}{c_7^3}. \tag{60}$$

The species (4, 7) *does not possess any more invariants. Therefore our species* (4, 7) *has been divided into the following three distinct sets:*

(A) *Those for which* $c_9\neq 0$. *The canonical form is* $y=x^{7/4}+x^{9/4}$.

(B) *Those for which* $c_9=0$ *but* $14c_7c_{13}-17c_{10}^2\neq 0$. *The canonical form is* $y=x^{7/4}+x^{13/4}$.

(C) *Those for which* $c_9=0$ *and* $14c_7c_{13}-17c_{10}^2=0$. *The canonical form is* $y=x^{7/4}$.

12. **Discussion of the species** (4, 6). It is clear that any curve of this species must be of the form

$$y = c_6x^{6/4} + c_8x^{8/4} + \cdots + c_{2p}x^{2p/4} + c_Qx^{Q/4} + \cdots, \tag{61}$$

where not only $c_6 \neq 0$ but also $c_Q \neq 0$ where $Q = 2\rho + 1 \geqq 7$ denotes the first *odd* power of the fourth root of x which appears in the series expansion for y. For otherwise our element would simply be an element of the species (2, 3), which shall be discussed at the end of our paper.

Upon substituting this into the equations (10) defining the group G', we find the parametric equations of the transformed element to be

$$(62)\qquad \begin{aligned} X &= \gamma_4 t^4 + \alpha_{10} c_6 t^6 + \gamma_8 t^8 + \gamma_{10} t^{10} + \cdots + \gamma_{2\rho} t^{2\rho} + \alpha_{10} c_{2\rho+1} t^{2\rho+1} + \cdots, \\ Y &= \beta_{10} c_6 t^6 + \delta_8 t^8 + \delta_{10} t^{10} + \cdots + \delta_{2\rho} t^{2\rho} + \beta_{10} c_{2\rho+1} t^{2\rho+1} + \cdots, \end{aligned}$$

where γ_4 and β_{10} are any two nonzero numbers and the remaining γ and δ are arbitrary numbers.

Let $X = T^4$. The first of the preceding equations defines t as an integral power series in T. This power series must be of the form

$$(63)\qquad t = A_1 T + A_3 T^3 + A_5 T^5 + \cdots + A_{2\rho-3} T^{2\rho-3} + A_{2\rho-2} T^{2\rho-2} + \cdots.$$

Until the $(2\rho - 3)$ term only odd powers appear. Thereafter the exponents of T increase by one (as far as we know).

Substituting this into the first of equations (62), we discover that all the $A_1 \neq 0$, A_3, $A_5, \cdots, A_{2\rho-3}$ are arbitrary but that we have $A_{2\rho-2} = c_{2\rho+1} A_1^{2\rho-5} A_3 / c_6$.

Finally replacing t by this series into the second of equations (62), we obtain the following transformation formulas between $(c_6, c_{2\rho+1}, c_{2\rho+3})$ and $(C_6, C_{2\rho+1}, C_{2\rho+3})$

$$(64)\qquad \begin{gathered} C_6 = \beta_{10} A_1^6 c_6, \qquad C_{2\rho+1} = \beta_{10} A_1^{2\rho+1} c_{2\rho+1}, \\ C_{2\rho+3} = \beta_{10}\left[(2\rho + 7) A_1^{2\rho} A_3 c_{2\rho+1} + A_1^{2\rho+3} c_{2\rho+3}\right]. \end{gathered}$$

From these results we immediately deduce that *the integer* $Q = 2\rho + 1 \geqq 7$ *is an arithmetic invariant and* c_Q^5 / c_6^{Q-1} *is a relative differential invariant*. In the remainder of this section, we shall prove that there are no more invariants.

First the transformation

$$(65)\qquad X = x + \frac{4 c_{2\rho+3}}{(2\rho + 7) c_6 c_{2\rho+1}}\, y, \qquad Y = y$$

will convert our arc (61) into an element of the form

$$(66)\qquad \begin{aligned} y = {} & d_6 x^{6/4} + d_8 x^{8/4} + \cdots + d_{2\rho} x^{2\rho/4} + d_{2\rho+1} x^{(2\rho+1)/4} \\ & + d_{2\rho+2} x^{(2\rho+2)/4} + d_{2\rho+4} x^{(2\rho+4)/4} + \cdots, \end{aligned}$$

where the d's are certain functions of the c's. We note that our transformation has eliminated the $(2\rho+3)$rd power of the fourth root of x.

Now we shall demonstrate the existence of a transformation which is the *inverse* of a correspondence of the form (37), which will carry our element (66)

into the canonical form $y=x^{3/2}+x^{Q/4}$. Thus the α_{0j} and the β_{ij} $(i=0, 1, 2)$ must be determined so that this correspondence (37) carries the canonical form into our element (66).

Now if we apply our transformation (37) to the canonical form $y=x^{6/4}+x^{(2p+1)/4}$, we find as the parametric form of the transformed element

$$
\begin{aligned}
X &= \alpha_{01}t^4 + \alpha_{02}t^8 + \alpha_{03}t^{12} + \cdots, \\
Y &= [\beta_{02}t^8 + (\beta_{03} + \beta_{20})t^{12} + \cdots + (\beta_{0,n+2} + \beta_{2,n-1})t^{4n+8} + \cdots] \\
&\quad + [\beta_{10}t^{2p+1} + \beta_{11}t^{2p+5} + \beta_{12}t^{2p+9} + \cdots + \beta_{1,p-2}t^{6p-7} \\
&\qquad + \beta_{1,p-1}t^{6p-3} + \beta_{1,p}t^{6p+1} + \cdots + \beta_{1,p-2+n}t^{6p-7+4n} + \cdots] \\
&\quad + [2\beta_{20}t^{2p+7} + 2\beta_{21}t^{2p+11} + \cdots + 2\beta_{2,p-2}t^{6p-1} \\
&\qquad + 2\beta_{2,p-1}t^{6p+3} + \cdots + 2\beta_{2,p-2+n}t^{6p-1+4n} + \cdots] \\
&\quad + [\beta_{10}t^6 + \beta_{11}t^{10} + \cdots + \beta_{1,p-2}t^{4p-2} + (\beta_{1,p-1} + \beta_{20})t^{4p+2} \\
&\qquad + (\beta_{1,p} + \beta_{21})t^{4p+6} + \cdots + (\beta_{1,p-2+n} + \beta_{2,n-1})t^{4p-2+4n} + \cdots].
\end{aligned}
\tag{67}
$$

In the second of the preceding equations, let us replace the coefficients of $t^6, t^{10}, t^{14}, \cdots, t^{4p-2}$ by $B_6, B_{10}, B_{14}, \cdots, B_{4p-2}$, respectively, the coefficient of $t^{6p-7+4n}$ by B_{6-7+4n}, the coefficient of $t^{2p+3+4n}$ by $B_{2p+3+4n}$, and the coefficient of t^{4n+4} by B_{4n+4} for $n=1, 2, 3, \cdots$. There obviously exists a one-to-one correspondence between the β_{ij} and the B's.

Therefore our problem is to determine the A's and the B's so that the equations

$$
\begin{aligned}
X &= T^4, \qquad t = A_1T + A_5T^5 + A_9T^9 + \cdots, \\
Y &= [B_8t^8 + B_{12}t^{12} + \cdots + B_{4n+4}t^{4n+4} + \cdots] \\
&\quad + [B_6(t^6 + t^{2p+1}) + B_{10}(t^{10} + t^{2p+5}) + \cdots + B_{4p-2}(t^{4p-2} + t^{6p-7})] \\
&\quad + [B_{6p-3}t^{6p-3} + B_{6p+1}t^{6p+1} + \cdots + B_{6p-7+4n}t^{6p-7+4n} + \cdots] \\
&\quad + [2B_{2p+7}t^{2p+7} + 2B_{2p+11}t^{2p+11} + \cdots + 2B_{6p-1}t^{6p-1} \\
&\qquad + 2B_{6p+3}t^{6p+3} + 2B_{6p+7}t^{6p+7} + \cdots + 2B_{6p-1+4n}t^{6p-1+4n} + \cdots] \\
&\quad + [(B_{6p-3} + B_{2p+7})t^{4p+2} + (B_{6p+1} + B_{2p+11})t^{4p+6} + \cdots \\
&\qquad + (B_{6p-7+4n} + B_{2p+3+4n})t^{4p-2+4n} + \cdots]
\end{aligned}
\tag{68}
$$

shall represent the element (66) upon elimination of the parameters t and T.

The elimination yields the following equivalence equations:

$$
\begin{aligned}
&B_8A_1^8 = d_8, \qquad B_{12}A_1^{12} + 8B_8A_1^7A_5 = d_{12}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots, \\
&\qquad B_{4n+4}A_1^{4n+4} + (\text{lower } B_{4k}) = d_{4n+4}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots;
\end{aligned}
\tag{69.1}
$$

$$
(69.2)\quad
\begin{aligned}
B_6A_1^6 &= d_6, & 6A_1^5A_5B_6 + B_{10}A_1^{10} &= d_{10},\\
B_6A_1^{2\rho+1} &= d_{2\rho+1}, & (2\rho+1)A_1^{2\rho}A_5B_6 + B_{10}A_1^{2\rho+5} &= d_{2\rho+5},\\
&\cdots\cdots\cdots\cdots\cdots\cdots,\\
6A_1^5A_{4\rho-7}B_6 + B_{4\rho-2}A_1^{4\rho-2} &+ (\text{lower } A_{4k-3} \text{ and } B_{4k-2}) = d_{4\rho+2},\\
(2\rho+1)A_1^{2\rho}A_{4\rho-7}B_6 + B_{4\rho-2}A_1^{6\rho-2} &+ (\text{lower } A_{4k-3} \text{ and } B_{4k-2}) = d_{6\rho-7};
\end{aligned}
$$

$$
(69.3)\quad
\begin{aligned}
(2\rho+1)A_1^{2\rho}A_{4\rho-3}B_6 + B_{6\rho-3}A_1^{6\rho-3} &+ (\text{lower } A_{4k-3} \text{ and } B_{2\rho-3+4k}) = d_{6\rho-3},\\
2B_{2\rho+7}A_1^7 &= d_{2\rho+7},\\
6A_1^5A_{4\rho-3}B_6 + (B_{6\rho-3} + B_{2\rho+7})A_1^{4\rho+2} &+ (\text{lower } A_{4k-3} \text{ and } B_{4k+2}) = d_{4\rho+2},\\
&\cdots\cdots\cdots\cdots\cdots\cdots,\\
(2\rho+1)A_1^{2\rho}A_{4\rho-7+4n}B_6 + B_{6\rho-7+4n}A_1^{6\rho-7+4n} &+ (\text{lower } A_{4k-3} \text{ and } B_{2\rho-3+4k})\\
&= d_{6\rho-7+4n},\\
2B_{2\rho+3+4n}A_1^{2\rho+3+4n} &+ (\text{lower } A_{4k-3} \text{ and } B_{2\rho+3+4k}) = d_{2\rho+3+4n},\\
6A_1^5A_{4\rho-7+4n}B_6 + (B_{6\rho-7+4n} + B_{2\rho+3+4n})A_1^{4\rho-6+4n} &+ \cdots = d_{4\rho-6+4n},\\
&\cdots\cdots\cdots\cdots\cdots\cdots.
\end{aligned}
$$

The first pair of equations (69.2) furnishes unique values for $A_1 \neq 0$ and $B_6 \neq 0$. The second pair yields unique values for A_5 and B_{10} since the determinant of the coefficients $(-2\rho+5)A_1^{2\rho+10}B_6 \neq 0$. The last pair give unique evaluations for $A_{4\rho-7}$ and $B_{4\rho-2}$ since the determinant of the coefficients $(-2\rho+5)A_1^{6\rho-2}B_6 \neq 0$. Thus the equations (69.2) determine the values of $A_1 \neq 0$, $A_5, \cdots, A_{4\rho-7}$; and $B_6 \neq 0$, $B_{10}, \cdots, B_{4\rho-2}$ uniquely.

The first triplet of the equations (69.3) determines unique values for $A_{4\rho-3}$, $B_{6\rho-3}$, and $B_{2\rho+7}$, since the determinant of the coefficients $2(2\rho-5)A_1^{8\rho+9}B_6 \neq 0$. The nth triplet furnishes unique evaluations for $A_{4\rho-7+4n}$, $B_{6\rho-7+4n}$, and $B_{2\rho+3+4n}$ since the determinant of the coefficients $2(2\rho-5)A_1^{8\rho+5+4n}B_6 \neq 0$. The equations (69.3) therefore give unique values for $A_{4\rho-3}$, $A_{4\rho+1}, \cdots, A_{4\rho-7+4n}, \cdots$; $B_{6\rho-3}$, $B_{6\rho+1}, \cdots, B_{6\rho-7+4n}, \cdots$; and $B_{2\rho+7}$, $B_{2\rho+11}, \cdots, B_{2\rho+3+4n}, \cdots$.

Finally the equations (69.1) obviously determine the values of B_8, $B_{12}, \cdots, B_{4n+4}, \cdots$ uniquely. Therefore this proves that *the canonical form of the species* (4, 6) *is* $y = x^{3/2} + x^{(2\rho+1)/4}$ *where* $\rho \geqq 3$.

We are now in a position to state the following result.

THEOREM 12. *The species* (4, 6) *possesses the odd integer* $Q \geqq 7$, *the first odd power of the fourth root of x which appears in the series expansion for y, as an arithmetic invariant, and the expression*

$$
(70)\qquad \frac{c_Q^6}{c_6^{Q-1}}
$$

as a relative differential invariant. There are no further invariants. The canonical form is $y=x^{3/2}+x^{Q/4}$.

13. **Discussion of the species** $(3, q)$. Any element of this species may be written in the form

$$y = c_q x^{q/3} + c_{q_1} x^{q_1/3} + c_{q_2} x^{q_2/3} + \cdots + c_{q_{n-1}} x^{q_{n-1}/3} + c_Q x^{Q/3} + \cdots, \tag{71}$$

where $q=3r+s$ $(0<s<3)$ and the integers $q<q_1<q_2< \cdots <q_{n-1}<Q$ are such that $q+q_i \not\equiv 0 \bmod 3$ $(i=1, 2, \cdots, n-1)$ but $q+Q\equiv 0 \bmod 3$.

The correspondent of this element under any transformation of our group G' (equations (12)) is given by the parametric form

$$\begin{aligned} X &= \alpha_{01}t^3 + \alpha_{02}t^6 + \cdots + \alpha_{0r}t^{3r} + \alpha_{10}c_{3r+s}t^{3r+s} + \cdots, \\ Y &= [\beta_{0,r+1}t^{3r+3} + \beta_{0,r+2}t^{3r+6} + \cdots] \\ &\quad + \beta_{10}[c_q t^q + c_{q_1}t^{q_1} + c_{q_2}t^{q_2} + \cdots + c_{q_{n-1}}t^{q_{n-1}} + c_Q t^Q + \cdots] \\ &\quad + \beta_{11}[c_q t^{q+3} + c_{q_1}t^{q_1+3} + c_{q_2}t^{q_2+3} + \cdots] \\ &\quad + \cdots. \end{aligned} \tag{72}$$

By the first of these equations, we find that t is defined as an integral power series in T where $X=T^3$. This power series must be of the form

$$t = A_1T + A_4T^4 + A_7T^7 + \cdots + A_{3r-2}T^{3r-2} + A_{3r-1}T^{3r-1} + \cdots, \tag{73}$$

where the exponents after the last written term increase by one (as far as we know). Substituting this into the first of equations (72), we discover that $(A_1\neq 0, A_4, A_7, \cdots, A_{3r-2}, A_{3r-1})$ may be taken as arbitrary quantities.

Now let us consider *the case where* $Q<2q-3$. In this case, it is found upon substituting (73) into the second of equations (72) that the following transformation formulas exist between (c_q, c_Q) and (C_q, C_Q):

$$C_q = \beta_{10}A_1^q c_q, \qquad C_Q = \beta_{10}A_1^Q c_Q. \tag{74}$$

Thus the *integer Q is an arithmetic invariant* (since $c_Q\neq 0$ and hence $C_Q\neq 0$), and *the expression c_Q^{q-1}/c_q^{Q-1} is a relative differential invariant.*

Next, by examining the cases $s=1$ and $s=2$ separately it may be shown that there exists a unique transformation which is the *inverse* of a correspondence of the form

$$\begin{aligned} X &= \alpha_{01}x + \alpha_{02}x^2 + \alpha_{03}x^3 + \cdots, \\ Y &= (\beta_{0,r+1}x^{r+1} + \beta_{0,r+2}x^{r+2} + \cdots) + y(\beta_{10} + \beta_{11}x + \beta_{12}x^2 + \cdots), \end{aligned} \tag{75}$$

which carries our element (71) for which $Q<2q-3$ into the normal form $y=x^{q/3}+x^{Q/3}$. *Therefore the canonical form of the species* $(3, q)$ *for which Q is the first power of the cube root of x in the series expansion for y such that* $q+Q\equiv 0 \bmod 3$ *and* $q<Q<2q-3$ *is* $y=x^{q/3}+x^{Q/3}$.

It remains now to examine *the case for which* $Q \geqq 2q-3$. First it may easily be verified that there exists a transformation which is the *inverse* of a correspondence of the form

$$X = x, \quad Y = (\beta_{r+1,0}x^{r+1} + \beta_{r+2,0}x^{r+2} + \cdots + \beta_{r+n,0}x^{r+n}) + y(1 + \beta_{11}x + \beta_{12}x^2 + \cdots + \beta_{1,r-1}x^{r-1}), \tag{76}$$

where $n=r-1$ or r according as $s=1$ or 2, which carries our element (71) into one of the form

$$y = d_q x^{q/3} + d_{2q-3}x^{(2q-3)/3} + \cdots, \tag{77}$$

where the d's are certain functions of the c's.

Next the point transformation

$$X = x + \frac{3d_{2q-3}}{qd_q^2}\,y, \qquad Y = y \tag{78}$$

carries our element (77) into one of the form

$$y = e_q x^{q/3} + e_{2q-2}x^{(2q-2)/3} + \cdots. \tag{79}$$

Finally it may very easily be shown by examining the cases $s=1$ and $s=2$ separately that there exists a transformation which is the *inverse* of the correspondence $X=x$, $Y=\psi(x, y)$ which carries our element (79) into the standard form $y=x^{q/3}$. *Therefore the canonical form of the species* (3, q) *for which* Q *is the first power of the cube root of* x *in the series expansion for* y *such that* $q+Q\not\equiv 0$ mod 3 *and* $Q\geqq 2q-3$ *is* $y=x^{q/3}$.

The following statement is thus found.

THEOREM 13. *In the species* (3, q) *let* Q *be the first power of the cube root of* x *in the series expansion for* y *such that* $q+Q\equiv 0$ mod 3. *If* Q *is such that* $q<Q<2q-3$, *then our species possesses* Q *as an arithmetic invariant and the expression*

$$\frac{c_Q^{q-1}}{c_q^{Q-1}} \tag{80}$$

as a relative differential invariant. For this case, there are no further relative invariants, the canonical form being $y=x^{q/3}+x^{Q/3}$. *On the other hand if* $Q\geqq 2q-3$, *there are no invariants and the canonical form is* $y=x^{q/3}$.

14. **Discussion of the species** (2, q). Obviously the inverse of the transformation

$$X = x, \quad Y = (c_{2r+2}x^{r+1} + c_{2r+4}x^{r+2} + \cdots) + y(c_{2r+1} + c_{2r+3}x + c_{2r+5}x^2 + \cdots) \tag{81}$$

carries any element of our species into the element $y=x^{(2r+1)/2}$. Hence

THEOREM 14. *The canonical form of the species* $(2, q)$ *is* $y=x^{q/2}$. *There are no differential invariants. The only arithmetic invariant is* q.

This completes our classification of analytic elements. This has been based on the type of lowest order differential invariant. (Higher invariants will be discussed elsewhere.) Accordingly our classification leads to *one regular type and twelve irregular types*. This is more complicated than the conformal classification discussed elsewhere(1). Projective invariants of irregular elements (besides Halphen's regular invariant) exist in great variety.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.
ILLINOIS INSTITUTE OF TECHNOLOGY,
CHICAGO, ILL.

THE TRANSFORMATION OF SERIES AND SEQUENCES

BY

W. T. SCOTT AND H. S. WALL

In this paper we present some results on certain aspects of the theory of summation of series and sequences. The paper is divided into three parts:

I. Linear manifolds of Hausdorff means.

II. Gronwall summability.

III. A method of summation arising from an algorithm of Schur.

The principal results of each part are summarized in the introductory paragraph of that part.

I. Linear manifolds of Hausdorff means

1. Introduction to Part I. The nth *Hausdorff mean* [7](1) of a series $\sum_{p=0}^{\infty} u_p$ is defined by

$$U_n = \sum_{p=0}^{n} C_{n,p}\Delta^{n-p}c_p\cdot(u_0 + u_1 + \cdots + u_p), \qquad n = 0, 1, 2, \cdots, \tag{1.1}$$

where $C_{n,p} = n!/p!(n-p)!$, $\Delta^i c_j = c_j - C_{i,1}c_{j+1} + C_{i,2}c_{j+2} - \cdots$, and $\{c_p\}$ is a *moment sequence*, i.e.,

$$c_p = \int_0^1 u^p d\phi(u), \qquad p = 0, 1, 2, \cdots, \phi(u) \in BV[0, 1](2). \tag{1.2}$$

The method of summation which assigns to the series $\sum u_p$ the sum s when $U_n \to s$ will be denoted by $[H, \phi(u)]$ or by $[H, c_p]$, and called a *Hausdorff mean.* $[H, \phi(u)]$ is *regular*(3) if and only if $\phi(u)$ is continuous at $u=0$ and $\phi(1)-\phi(0)=1$. We shall say that $[H, \phi(u)]$ is *essentially regular* if $\phi(u)$ is continuous at $u=0$. In this case, if $\phi(1)-\phi(0)=c_0\neq 0$, the mean $[H, \phi(u)/c_0]$ is regular.

It will be observed that if in (1.2) the integrand u^p is replaced by some other suitably chosen function, e.g., u^{p+1}, then no restriction need be placed upon $\phi(u) \in BV[0, 1]$ in order that the resulting mean be essentially regular.

Presented to the Society in three parts: *Linear manifolds of Hausdorff means*, November 22, 1940; *Gronwall summability*, April 12, 1941; *The transformation of series and sequences*, April 12, 1941; received by the editors January 23, 1941, and, in revised form, February 28, 1941. We are grateful to Professor Einar Hille for directing our attention to the possibility of connecting our investigations with the ideas of Gronwall.

(1) Numbers in square brackets refer to bibliography at end of paper.

(2) $BV[a, b]$ is the class of all functions, real and complex, of a real variable u, which are of bounded variation on $a \leqq u \leqq b$.

(3) That is, it assigns to any convergent series $\sum u_n$ the value $\sum u_n$.

In the light of this observation, we consider the problem of determining conditions upon a sequence of functions $\{\beta_p(u)\}$ in order that(4)

$$(1.3) \qquad c_p = \int_0^1 \beta_p(u)d\phi(u), \qquad p = 0, 1, 2, \cdots,$$

be a moment sequence for every $\phi(u)$ in $BV[0, 1]$; and of determining further conditions under which $[H, c_p]$ is essentially regular. The results are contained in Theorems 2.1–2.4. We call the set of all means obtained with a given sequence $\{\beta_p(u)\}$ *a (linear) manifold*, denote it by $M[\beta_p(u)]$, and call $\{\beta_p(u)\}$ the *basis* of the manifold.

Perhaps the chief interest in this theory lies in the fact that we are able to obtain infinite classes of means all including a given mean or all equivalent to a given mean. For instance, if in (1.3) $\beta_p(u) = 1/(1+pu)$ and $\phi(u)$ is monotone non-decreasing, $\phi(1)-\phi(0)=1$, then $[H, c_p]$ is a regular mean included in $(C, 1)$; is equivalent to $(C, 1)$ if and only if $\int_0^1 d\phi(u)/u$ converges; and is equivalent to $(C, 0)$ (convergence) if and only if $\phi(u)$ is discontinuous at $u=1$.

2. **Conditions for a basis.** One readily sees that *necessary* conditions upon a sequence of functions $\{\beta_p(u)\}$ in order that it be a basis for a manifold $M[\beta_p(u)]$ are:

(a) $\beta_p(u)$ *is continuous for* $0 \leq u \leq 1$;

(b) $\{\beta_p(u)\}$ *is a moment sequence for each fixed* u, $0 \leq u \leq 1$.

The first of these conditions is necessary in order that the Stieltjes integral $\int_0^1 \beta_p(u)d\phi(u)$ exist for all $\phi(u)$ in $BV[0, 1]$. The second is necessary in order that the sequence of these integrals be a moment sequence when $\phi(u)$ is a step function with a single point of increase. From (b) it follows that there must exist a function(5) $B(u, t)$ of bounded variation in t for each u, $0 \leq u \leq 1$, such that

$$(2.1) \qquad \beta_n(u) = \int_0^1 t^n d_t B(u, t), \qquad n = 0, 1, 2, \cdots.$$

If (a), (b) hold, then a *sufficient* condition for $\{\beta_p(u)\}$ to be a basis is that

$$(2.2) \qquad \sum_{p=0}^{n} C_{n,p} \left| \Delta^{n-p}\beta_p(u) \right| < M, \qquad M \text{ independent of } n, u.$$

In fact, if

$$(2.3) \qquad a_n = \int_0^1 \beta_n(u)d\alpha(u), \qquad \alpha(u) \in BV[0, 1],$$

then, when (2.2) holds,

(4) Throughout this paper the integrals are in the Riemann-Stieltjes sense.

(5) This function is uniquely determined to an additive constant for all t where it is continuous [7].

$$\sum_{p=0}^{n} C_{n,p} \left| \Delta^{n-p} a_p \right| < MV,$$

where V is the total variation of $\alpha(u)$ on the interval $0 \leqq u \leqq 1$, and consequently $\{a_p\}$ is a moment sequence by virtue of a theorem of Hausdorff [7]. Condition (2.2) is met, in particular, if $\{\beta_p(u)\}$ is a totally monotone sequence([6]) for each u, $0 \leqq u \leqq 1$. In this case we shall call $\{\beta_p(u)\}$ a *totally monotone basis.*

THEOREM 2.1. *If $B(u, t)$ is real and of bounded variation in t, $0 \leqq t \leqq 1$, uniformly for all u in the interval $0 \leqq u \leqq 1$, and if the functions $\beta_n(u)$ given by* (2.1) *are continuous functions of u, $0 \leqq u \leqq 1$, then $\{\beta_p(u)\}$ is a basis for a manifold $M[\beta_p(u)]$.*

REMARK. A *sufficient* condition [3] for the continuity of the functions $\beta_p(u)$ is that $B(u, t)$ be of bounded variation in t uniformly for all u, and be continuous in u for an everywhere dense set of values of t including $t = 0, 1$. This everywhere dense set may depend upon u.

To prove the theorem, put $P(u, t) = \frac{1}{2}(\int_0^t |d_t B(u, t)| + \int_0^t d_t B(u, t))$, $N(u, t) = \frac{1}{2}(\int_0^t |d_t B(u, t)| - \int_0^t d_t B(u, t))$. It is no restriction to assume that $B(u, 0) = 0$. Then $B(u, t) = P(u, t) - N(u, t)$, and $P(u, t)$, $N(u, t)$ are monotone non-decreasing functions of t for each fixed u. We then have

$$S_n = \sum_{p=0}^{n} C_{n,p} \left| \Delta^{n-p} \beta_p(u) \right| = \sum_{p=0}^{n} C_{n,p} \left| \Delta^{n-p} \int_0^1 t^p d_t B(u, t) \right|$$
$$\leqq \sum_{p=0}^{n} C_{n,p} \left| \Delta^{n-p} \pi_p(u) - \Delta^{n-p} \nu_p(u) \right|,$$

where $\pi_p(u) = \int_0^1 t^p d_t P(u, t)$, $\nu_p(u) = \int_0^1 t^p d_t N(u, t)$. Hence,

$$S_n \leqq \sum_{p=0}^{n} C_{n,p} \Delta^{n-p} \pi_p(u) + \sum_{p=0}^{n} C_{n,p} \Delta^{n-p} \nu_p(u)$$

or

$$S_n \leqq \pi_0(u) + \nu_0(u) = T(u),$$

where $T(u)$ is the total variation of $B(u, t)$ in the interval $0 \leqq t \leqq 1$. Since, by hypothesis, $T(u) < M$, where M is independent of u, we see that (2.2) holds, and the theorem is thereby established.

DEFINITION. *A manifold is called regular if it contains at least one regular mean, and if every mean contained in it is essentially regular.*

THEOREM 2.2. *A totally monotone basis $\{\beta_p(u)\}$ is the basis of a regular manifold if and only if $\beta_0(u) \not\equiv 0$, and*

([6]) A real sequence $\{c_n\}$ is totally monotone if all differences $\Delta^m c_n \geqq 0$.

$$(2.4)\qquad \lim_{n=\infty} \Delta^n\beta_0(u) = 0, \qquad 0 \leq u \leq 1(^7).$$

Proof. The condition $\beta_0(u)\not\equiv 0$ is obviously necessary. Let $[H, a_p]$ be any mean in $M[\beta_p(u)]$. Then $[H, a_p]$ is essentially regular if and only if [7] $\lim_{n=\infty} \Delta^n a_0=0$. It readily follows that the condition (2.4) is *necessary* in order that $M[\beta_p(u)]$ be regular. The conditions are also *sufficient*. For, since $\{\beta_p(u)\}$ is totally monotone, the sequence $\{\Delta^p\beta_0(u)\}$ is monotone non-increasing, and therefore $\lim_{n=\infty} \Delta^n a_0=\lim_{n=\infty} \int_0^1\Delta^n\beta_0(u)d\alpha(u)=0$, so that $[H, a_p]$ is essentially regular when (2.4) holds. Since $\beta_0(u)\not\equiv 0$ the manifold must contain at least one regular mean.

THEOREM 2.3. *Let* $\{\beta_p(u)\}$, $\beta_0(u)\not\equiv 0$, *be a basis given by* (2.1), *where* $B(u, t)$ *is of bounded variation in* t, $0\leq t\leq 1$, *for each* u, $0\leq u\leq 1$. *Let* $|B(u, t)|<K$, $0\leq t\leq 1$, $0\leq u\leq 1$, *where* K *is a finite constant independent of* u *and* t; *and let* $\lim_{t=0+} B(u, t)=B(u, 0)=0$ *uniformly for* $0\leq u\leq 1$. *Then* $M[\beta_p(u)]$ *is a regular manifold.*

Proof. Let a_n be given by (2.3). Then we shall prove that $\lim_{n=\infty} \Delta^n a_0=0$. We have

$$\Delta^n a_0 = \int_0^1 \Delta^n\beta_0(u)d\alpha(u) = \int_0^1\int_0^1 (1-t)^n d_tB(u, t)d\alpha(u).$$

Denote the inner integral by I. Then $I=I_1+I_2$ where $I_1=\int_0^s(1-t)^n d_tB(u, t)$, $I_2=\int_s^1(1-t)^n d_tB(u, t)$. After an integration by parts in I_1 we get

$$I_1 = (1-s)^nB(u, s) - \int_0^s B(u, t)d_t[(1-t)^n].$$

Consequently,

$$\begin{aligned} |I_1| &\leq (1-s)^n|B(u, s)| + \operatorname*{l.u.b.}_{0\leq t\leq s} |B(u, t)| \int_0^s |d_t(1-t)^n| \\ &\leq (1-s)^n\epsilon + [1-(1-s)^n]\epsilon = \epsilon, \qquad \epsilon > 0, \end{aligned}$$

for all u, $0\leq u\leq 1$, provided only that $s<s_0$ where s_0 is sufficiently small.

Having chosen $s<s_0$, we integrate by parts in I_2 and get

$$I_2 = -(1-s)^nB(u, s) - \int_s^1 B(u, t)d_t(1-t)^n,$$

$$\begin{aligned} |I_2| &\leq (1-s)^n|B(u, s)| + \operatorname*{l.u.b.}_{s\leq t\leq 1} |B(u, t)| \int_s^1 |d_t(1-t)^n| \\ &\leq (1-s)^n\epsilon + K(1-s)^n = (1-s)^n(K+\epsilon). \end{aligned}$$

Hence, if n_0 is sufficiently large, $|I_2|<\epsilon$ if $n>n_0$.

(7) This theorem holds for any basis such that $\{\Delta^m\beta_0(u)\}$ is uniformly bounded on $0\leq u\leq 1$.

We then have

$$|\Delta^n a_0| \leqq \int_0^1 (|I_1| + |I_2|)\,|d\alpha(u)| \leqq 2\epsilon T, \qquad n > n_0,$$

where T is the total variation of $\alpha(u)$ in $0 \leqq u \leqq 1$. Thus $\lim_{n\to\infty} \Delta^n a_0 = 0$, as was to be proved. Hence, every mean in $M[\beta_p(u)]$ is essentially regular; and since $\beta_0(u) \not\equiv 0$ it must contain at least one regular mean.

We now have the following *comparison theorem*.

THEOREM 2.4. *Let* $\{\beta_p(u)\}$ *be a sequence of continuous real functions on the interval* $0 \leqq u \leqq 1$, $\beta_0(u) \not\equiv 0$. *Then* $\{\beta_p(u)\}$ *is a basis if there exists a totally monotone basis* $\{\pi_p(u)\}$ *such that* $\Delta^m \beta_n(u) \leqq \Delta^m \pi_n(u)$, $0 \leqq u \leqq 1$, $m, n = 0, 1, 2, \cdots$. *Moreover, if* $M[\pi_p(u)]$ *is regular, and* $\lim_{n\to\infty} \Delta^n \beta_0(u) = 0$, $0 \leqq u \leqq 1$, *then* $M[\beta_p(u)]$ *is regular.*

Proof. Put $\alpha_p(u) = \pi_p(u) - \beta_p(u)$. Then $\Delta^n \alpha_p(u) = \Delta^n \pi_p(u) - \Delta^n \beta_p(u) \geqq 0$, so that $\{\alpha_p(u)\}$ is a totally monotone basis. Thus, $\beta_p(u) = \pi_p(u) - \alpha_p(u)$ where $\{\pi_p(u)\}$ and $\{\alpha_p(u)\}$ are totally monotone bases. It readily follows that $\{\beta_p(u)\}$ is a basis. If $M[\pi_p(u)]$ is regular, so that $\lim_{n\to\infty} \Delta^n \pi_0(u) = 0$, and if $\lim_{n\to\infty} \Delta^n \beta_0(u) = 0$, then it follows that $\lim_{n\to\infty} \Delta^n \alpha_0(u) = 0$. Hence if $a_n = \int_0^1 \beta_n(u) d\phi(u)$, $\phi(u) \in BV[0, 1]$, we must have $\lim_{n\to\infty} \Delta^n a_0 = 0$, so that every mean in $M[\beta_p(u)]$ is essentially regular. The condition $\beta_0(u) \not\equiv 0$ insures that $M[\beta_p(u)]$ contains at least one regular mean. This completes the proof of the theorem.

DEFINITION. *A manifold* $M[\beta_p(u)]$ *is said to include a given regular mean* $[H, b_p]$ *if any series which* $[H, b_p]$ *sums is also summed by every regular mean in* $M[\beta_p(u)]$.

THEOREM 2.5. *Let* $M[\beta_p(u)]$ *be regular, and* $[H, b_p]$ *a regular mean for which* $b_p \neq 0$, $p = 0, 1, 2, \cdots$. *Then* $M[\beta_p(u)]$ *includes* $[H, b_p]$ *if and only if* $\{\beta_p(u)/b_p\}$ *is the basis for a regular manifold.*

This follows at once from the theorem of Hausdorff [7] that a regular mean $[H, a_p]$ includes a regular mean $[H, b_p]$ for which $b_p \neq 0$, $p = 0, 1, 2, \cdots$, if and only if $\{a_p/b_p\}$ is a moment sequence and $[H, a_p/b_p]$ a regular mean.

ILLUSTRATIVE EXAMPLES. Let $\beta_n^{(k)}(u) = (u+1)/[u+(n+1)^k]$, $n = 0, 1, 2, \cdots$, where k is a positive integer. The sequence is totally monotone when $k = 1$, and $\beta_n^{(1)}(u) = \int_0^1 t^n d_t B_1(u, t)$, $B_1(u, t) = t^{u+1}$. By Theorem 2.2, this is the basis of a regular manifold M_1.

When $k = 2$, we find $B_2(u, t) = t[\cos(u^{1/2} \log t) - u^{-1/2} \sin(u^{1/2} \log t)]$, so that the sequence is not totally monotone. Since $d_t B_2(u, t) = -u^{-1/2}(1+u) \cdot \sin(u^{1/2} \log t)\, dt$, it follows that $\int_0^1 |d_t B_2(u, t)| \leqq 2$, $0 \leqq u \leqq 1$, $0 \leqq t \leqq 1$. Also, $|B_2(u, t)| \leqq t(1 - \log t)$, $0 \leqq t \leqq 1$. It therefore follows from Theorems 2.1, 2.3 that $M[\beta_p^{(2)}(u)]$ is a regular manifold.

On writing $\beta_p^{(k)}(u)$ as a sum of partial fractions it is now easy to see that $\{\beta_p^{(k)}(u)\}$ is a basis for a regular manifold M_k, $k=1, 2, 3, \cdots$.

To illustrate Theorem 2.5, we shall show that M_1 includes $(C, 1)$ but does not include (C, k) if $k>1$. The sequence $\{\beta_n^{(1)}(u)/b_n\}$, $b_n=1/(1+n)$, must be proved to be a basis for a regular manifold. We find that

$$\beta_n^{(1)}(u)/b_n = (u+1)(n+1)/(u+n+1) = \int_0^1 t^n d_tQ(u, t),$$

where $Q(u, t)=-ut^{u+1}$, $0\leqq t<1$, and $Q(u, 1)=1$. By Theorem 2.3, this is the basis of a regular manifold, so that M_1 includes $(C, 1)$. On the other hand, if $b_n=(1+n)^{-k}$, $k>1$, then $\beta_n^{(1)}(u)/b_n$ tends to ∞ with n, so that M_1 cannot include (C, k), $k>1$.

We shall give an elementary proof that M_k includes (C, k) but not $(C, k+\theta)$, $\theta>0$. An arbitrary mean in M_k has the form $[H, b_n]$ where

$$(2.5)\qquad \begin{aligned} b_n &= \left[\frac{c_0}{(n+1)^k} + \frac{(n+1)^k-1}{(n+1)^k}\left(\frac{c_1}{(n+1)^k} - \frac{c_2}{(n+1)^{2k}} + \cdots\right)\right], \\ c_n &= \int_0^1 u^n d\phi(u), \qquad \phi(u) \in BV[0, 1]. \end{aligned}$$

We observe that the mean $[H, b_n^*]$ obtained by replacing c_n by c_{n+1}, $n=0, 1, 2, \cdots$, is also in M_k. This amounts to using $ud\phi(u)$ instead of $d\phi(u)$. Of course $[H, b_n]$ is regular if and only if $b_0=1$, i.e., $c_0=1$.

To prove that $[H, b_n]\supset(C, k)$ we must show that the sequence $\{(n+1)^kb_n\}$ is a regular sequence[8] when $c_0=1$. We have

$$(n+1)^kb_n=c_0+c_1-\left[\frac{c_1}{(n+1)^k}+\frac{(n+1)^k-1}{(n+1)^k}\left(\frac{c_2}{(n+1)^k}-\frac{c_3}{(n+1)^{2k}}+\cdots\right)\right].$$

On comparing this with (2.5) we see that $(n+1)^kb_n=r\cdot 1+s\cdot b_n^*$, $r+s=1$, is a linear combination of regular sequences, where the constants of combination add up to unity, and is therefore a regular sequence. Hence we have proved that $[H, b_n]\supset(C, k)$. Since, in general, $(n+1)^{k+\theta}b_n\to\infty$ as $n\to\infty$ if $\theta>0$, it follows that $[H, b_n]$ does not include $(C, k+\theta)$.

3. **The manifold $M[(1+pu)^{-1}]$.** The sequence $\mu_p=(1+pu)^{-1}$, $p=0, 1, 2, \cdots$, is a totally monotone basis inasmuch as $\mu_p(u)$ is continuous and $\Delta^m\mu_n\geqq 0$, $m, n=0, 1, 2, \cdots$, $0\leqq u\leqq 1$. Since $\lim_{n\to\infty}\Delta^n\mu_0=0$, $0\leqq u\leqq 1$, the manifold $M[\mu_p(u)]$ is regular, by Theorem 2.2. This can be established also by Theorem 2.3. For, if $M(u, t)$ is the function defined as follows:

$$(3.1)\qquad M(u, t) = \begin{cases} t^{1/u}, & 0<u\leqq 1, \quad 0\leqq t\leqq 1, \\ 0, & u=0, \quad 0\leqq t<1, \\ 1, & u=0, \quad t=1, \end{cases}$$

[8] A regular sequence is one which determines a regular mean.

then $\mu_p(u)=\int_0^1 t^p d_t M(u, t)$, $p=0, 1, 2, \cdots$. It is readily seen that (3.1) satisfies the conditions imposed upon $B(u, t)$ in Theorem 2.3.

We shall begin by determining a large class of means which are in this manifold. Let

$$a_n = \int_0^1 \frac{d\alpha(u)}{1+nu}, \qquad n = 0, 1, 2, \cdots, \alpha(u) \in BV[0, 1], \tag{3.2}$$

so that $[H, a_n]$ is a mean in $M[\mu_p(u)]$. Since $\{a_n\}$ is a moment sequence, there must exist a function $\theta(t)$ in $BV[0, 1]$ such that

$$a_n = \int_0^1 t^n d\theta(t), \qquad n = 0, 1, 2, \cdots. \tag{3.3}$$

One may determine $\theta(t)$ in the following way. Put $p(x)=a_0-a_1x+a_2x^2-\cdots$. Then, on putting in the values of the a_n's from (3.2) and using a well known integral representation for hypergeometric functions, we get

$$p(x) = \int_0^1 \frac{d\theta(t)}{1+xt} = \int_0^1 \int_0^1 \frac{d_t M(u, t)}{1+xt} d\alpha(u) = \int_0^1 \frac{d_t \int_0^1 M(u, t) d\alpha(u)}{1+xt},$$

where $M(u, t)$ is given by (3.1), and where the last step may be verified directly from the definition of a Stieltjes integral. Consequently,

$$\theta(t) = \int_0^1 M(u, t) d\alpha(u), \qquad 0 \leqq t \leqq 1, \tag{3.4}$$

We shall now prove the following theorem.

THEOREM 3.1. *Let $\theta(t)=p_1t+p_2t^2+p_3t^3+\cdots$ be any power series with constant term* 0 *and which is absolutely convergent when $t=1$. Then $\theta(t) \in BV[0, 1]$, and $[H, \theta(t)]$ is a mean in $M[(1+pu)^{-1}]$.*

Proof. Let

$$\alpha(u) = \begin{cases} p_1+p_2+p_3+p_4+\cdots, & 1 \leqq u, \\ p_2+p_3+p_4+\cdots, & 1/2 \leqq u < 1, \\ p_3+p_4+\cdots, & 1/3 \leqq u < 1/2, \\ \cdots, & \cdots, \\ 0, & u = 0. \end{cases}$$

Then $\alpha(u) \in BV[0, 1]$; and if this function is used in (3.2) there is determined a mean $[H, a_n]$ in $M[\mu_p(u)]$. We see by (3.3), (3.4) that $[H, a_n]=[H, \theta(t)]$ where $\theta(t)=p_1t+p_2t^2+\cdots$. This proves the theorem.

The function $1-(1-t)^\alpha$, $\Re(\alpha)>0$, which is the mass function for Cesàro summability (C, α), has a power series of the kind specified in Theorem 3.1, and hence (C, α) is in the manifold $M[\mu_p(u)]$.

Throughout the remainder of this section we shall consider only the subset of means of $M[\mu_p(u)]$ of the form $[H, g(p)]$ where

$$(3.5)\qquad g(x) = \int_0^1 \frac{d\phi(u)}{1 + xu}, \qquad \phi(u) \text{ monotone, } \phi(1) - \phi(0) = 1.$$

We recall that $g(x)$ has a continued fraction representation of the form [11]

$$(3.6)\qquad g(x) = \frac{1}{1} + \frac{g_1 x}{1} + \frac{(1 - g_1)g_2 x}{1} + \frac{(1 - g_2)g_3 x}{1} + \cdots,$$

where $0 \leqq g_n \leqq 1$, $n = 1, 2, 3, \cdots$, it being agreed that if some g_n is 0 or 1 the continued fraction terminates with the first identically vanishing partial quotient. Conversely, any such continued fraction represents a function of the form (3.5). The function $g(x)$ is holomorphic in the plane of x cut along the real axis from $x = -1$ to $x = -\infty$.

THEOREM 3.2. $(C, 0) \subset [H, g(p)] \subset (C, 1)$.

Proof. Since $[H, g(p)]$ is regular, it includes $(C, 0)$ (convergence). To prove the second relation, put $g^*(x) = 1/1 + (1-g_1)x/1 + g_1(1-g_2)x/1 + \cdots$, the continued fraction being obtained from (3.6) by replacing g_n by $1 - g_n$, $n = 1, 2, 3, \cdots$. Then, we have the continued fraction identity [11, p. 166]

$$(3.7)\qquad g(x)g^*(x) = 1/(1 + x), \quad \text{all } x \text{ not real and } \leqq -1.$$

Since $[H, g^*(p)]$ is a regular mean, it follows that $[1/(1+p)]: g(p)$, $p = 0, 1, 2, \cdots$, is a regular sequence, so that, since $(C, 1) = [H, 1/(1+p)]$, the inclusion relation follows by Hausdorff's fundamental theorem.

It is perhaps of interest to determine conditions under which $[H, g(p)] \approx (C, 0)$, $[H, g(p)] \approx (C, 1)$. We have this theorem.

THEOREM 3.3. *If $g(x)$ is the function* (3.5), *then* $[H, g(p)] \approx (C, 0)$ *if and only if $\phi(u)$ is discontinuous at $u = 0$; and* $[H, g(p)] \approx (C, 1)$ *if and only if $\int_0^1 d\phi(u)/u$ is finite.*

Proof. From the equation

$$(n + 1)\int_0^1 \frac{d\phi(u)}{1 + nu} = \left(1 + \frac{1}{n}\right)\int_0^1 \frac{d\phi(u)}{(1/n) + u}, \qquad n > 0,$$

it follows that $\{(n+1)g(n)\}$ is a bounded sequence if and only if $\int_0^1 d\phi(u)/u$ converges. Moreover, when $\{(n+1)g(n)\}$ is bounded, it is a regular moment sequence. For, in terms of the bounded monotone function $\psi(u) = \int_0^u (1-t)d\phi(t)/t$, we have

$$(n + 1)g(n) = \int_0^1 \frac{d\phi(u)}{u} - \int_0^1 \frac{d\psi(u)}{1 + nu},$$

and thus have expressed our sequence as a linear combination of two regular sequences, where the constants of combination add up to unity. Hence we have shown that $g(p):(1+p)^{-1}$, $p=0, 1, 2, \cdots$, is a regular sequence, i.e., $[H, g(p)]\supset(C, 1)$, if and only if $\int_0^1 d\phi(u)/u<\infty$. Since we previously had $[H, g(p)]\subset(C, 1)$, the second part of the theorem is proved

To prove the first part, we use (3.7) and write $1/g(p)=(p+1)g^*(p)$. Hence, from the foregoing proof, $\{1/g(p)\}$ is a regular sequence if (and only if) it is bounded, i.e., if and only if

$$\lim_{n=\infty} g(n) = \lim_{n=\infty} \int_0^1 \frac{d\phi(u)}{1+nu} > 0.$$

But this well known limit is equal to $\phi(0+)-\phi(0)$, the discontinuity of $\phi(u)$ at $u=0$. This completes the proof of the theorem.

From the above proof we see that an alternative form of Theorem 3.3 is as follows.

THEOREM 3.4. *If $g(x)$ is the function* (3.5), *then* $[H, g(p)]\approx(C, 0)$ *if and only if the sequence* $\{1/g(p)\}$ *is bounded; and* $[H, g(p)]\approx(C, 1)$ *if and only if the sequence* $\{(p+1)g(p)\}$ *is bounded.*

The conditions may be formulated in terms of the continued fraction (3.6) for $g(x)$ as follows.

THEOREM 3.5. *If $g(x)$ is the function* (3.5) *with continued fraction representation* (3.6), *then* $[H, g(p)]\approx(C, 0)$ *if and only if the series obtained from*

$$1+\sum_{p=1}^{\infty} \frac{g_1g_2\cdots g_p}{(1-g_1)(1-g_2)\cdots(1-g_p)} \tag{3.8}$$

by replacing g_{2k} by $1-g_{2k}$, $k=1, 2, 3, \cdots$, converges; and $[H, g(p)]\approx(C, 1)$ *if and only if the series obtained from* (3.8) *by replacing g_{2k-1} by $1-g_{2k-1}$, $k=1, 2, 3, \cdots$, converges.*

Proof. Let $G(x)$, $H(x)$ be the functions obtained from (3.6) by substituting $1-g_{2k}$ for g_{2k} and $1-g_{2k-1}$ for g_{2k-1}, respectively, $k=1, 2, 3, \cdots$. Then [12] $G(-x[1+x]^{-1})=1/g(x)$, $H(-x[1+x]^{-1})=(1+x)g(x)$. As $x\to+\infty$, we see that $-x/(1+x)\to-1^+$. The conditions stated are necessary and sufficient [11, p. 181] for $G(x)$ and $H(x)$, respectively, to remain finite as $x\to-1$ through real values greater than -1. Hence the theorem follows by Theorem 3.4.

EXAMPLES. If $g_p=1/2$, $p=1, 2, 3, \cdots$, in (3.6), then $g(p)=(1+p)^{-1/2}$, and therefore $[H, g(p)]=[H, 1/2]\approx(C, 1/2)$. If $g_{2k-1}=1/2$, $g_{2k}=2/3$ then $[H, g(p)]\approx(C, 0)$; while if $g_{2k-1}=2/3$, $g_{2k}=1/2$ then

$$[H, g(p)] \approx (C, 1).$$

If the coefficients $p_1, p_2, p_3, \cdots$ in the power series of Theorem 3.1 are

real and greater than or equal to 0, and $\sum p_n = 1$, then $[H, \theta(t)]$ is a mean of the kind being considered. In particular, if $\theta(t) = 1-(1-t)^\alpha$, $0 \leqq \alpha \leqq 1$, the coefficients are greater than or equal to 0 and their sum is unity, and consequently the Cesàro means (C, α), $0 \leqq \alpha \leqq 1$, are all of the form $[H, g(p)]$ where $g(x)$ is of the form (3.5).

4. **Inclusion problems in the difference matrix for $\{g(p)\}$.** If $\{c_n\}$ is a moment sequence, then the row, column, and diagonal sequences in the difference matrix [5]

$$(\Delta^m c_n) = \begin{pmatrix} c_0, & c_1, & c_2, & \cdots \\ \Delta c_0, & \Delta c_1, & \Delta c_2, & \cdots \\ \Delta^2 c_0, & \Delta^2 c_1, & \Delta^2 c_2, & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{pmatrix}$$

are all moment sequences, and accordingly define Hausdorff means. The following theorem concerns the rows in this matrix when the base sequence $\{c_p\}$ is of the form $\{g(p)\}$. Its proof affords an application of some of the ideas of §2.

THEOREM 4.1. *Let $g(x)$ be a function of the form*

$$g(x) = \int_0^1 \frac{d\phi(u)}{1 + xu}, \qquad \phi(u) \in BV[0, 1]. \tag{4.1}$$

Put $b_n = \Delta^m g(n)/\Delta^m g(0)$, supposing $\Delta^k g(0) \neq 0$, $k = 0, 1, 2, \cdots$. Then $[H, b_n]$ is a regular mean, and

$$[H, b_n] \supset (C, m), \qquad m = 1, 2, 3, \cdots. \tag{4.2}$$

Proof. Since $\lim_{n\to\infty} \Delta^n g(0) = 0$ it follows that $\lim_{n\to\infty} \Delta^n b_0 = 0$. Thus $[H, b_n]$ is regular. We use Theorems 2.5, 2.3 to prove (4.2). To do this, we note that $\{\beta_p(u)\}$, where

$$\beta_p(u) = \Delta^m(1 + pu)^{-1} = m!u^m/(1 + pu)(1 + [p + 1]u) \cdots (1 + [p + m]u),$$

is a totally monotone basis for a regular manifold, $M[\beta_p(u)]$, and that the mean $[H, b_n]$ is in $M[\beta_p(u)]$. To prove (4.2) it is sufficient, by Theorem 2.5, to prove that the sequence $\{\beta_p(u)/c_p\}$, where $c_p = C_{p+m,m}$, $p = 0, 1, 2, \cdots$, is the moment sequence defining (C, m), is the basis for a regular manifold; and this will in turn be true if

$$\gamma_n(u) = C_{n+m}\beta_n(u) = \int_0^1 t^n d_t B(u, t), \qquad n = 0, 1, 2, \cdots, \tag{4.3}$$

where $B(u, t)$ satisfies the conditions of Theorem 2.3.

Since $\gamma_n(0) = 0$, we must have $B(0, t) \equiv 0$, $0 \leqq t \leqq 1$.

If $0 < u \leqq 1$, we find that

$$(4.4)\qquad \gamma_n(u) = \frac{A_0}{1+nu} + \frac{A_1}{1+(n+1)u} + \cdots + \frac{A_m}{1+(n+m)u},$$

where $u^m A_k$ is a polynomial in u of degree m with coefficients independent of n. Hence it follows that if $u>0$, $\gamma_n(u)$ is a linear combination of totally monotone sequences with coefficients of combination independent of n. Therefore $\{\gamma_n(u)\}$ is a moment sequence. Hence there exists a function $B(u, t)$ in $BV[0, 1]$ such that (4.3) holds, $0 \leqq u \leqq 1$.

We show next that

$$(4.5)\qquad \lim_{t=0^+} B(u, t) = 0, \qquad \text{uniformly for } 0 \leqq u \leqq 1.$$

If $u>0$, it follows from (4.4) that $B(u, t)$ can be expressed as the sum of $m+1$ functions of the form $[h(u)t^p t^{1/u}]/u^m$, p an integer, $0 \leqq p \leqq m$, where $h(u)$ is a polynomial in u. Hence it is easy to see that $|B(u, t)| < kt$, where k is a constant independent of u, t, provided that $0<t\leqq r<1$, r a fixed number less than 1, $0<u\leqq 1$. It follows that (4.5) holds.

It remains to be shown that

$$(4.6)\qquad |B(u, t)| \leqq K, \quad 0 \leqq u \leqq 1, 0 \leqq t \leqq 1, K < \infty \text{ and independent of } u, t.$$

The proof is by induction starting with $m=0$, for which the requirement is clearly met. If $u>0$, $\{\gamma_n(u)f(m, u)\}$ is a regular sequence if $f(m, u)$ is chosen so that $\gamma_0 f(m, u)=1$. We shall use the superscript m to indicate dependence upon m.

To prove (4.6) by induction, we assume that

$$|B^{(m)}(u, t)| \leqq K^{(m)}, \qquad 0 \leqq u \leqq 1, 0 \leqq t \leqq 1, m \geqq 0,$$

and shall prove it for $m+1$. We have

$$\gamma_n^{(m+1)}(u)f(m+1, u) = \frac{(n+m+1)u}{1+(n+m+1)u}\, e(m, u)\cdot\gamma_n^{(m)}(u)f(m, u),$$

where $e(m, u) = [1+(m+1)u]/(m+1)u$. Then, we may write

$$\frac{(n+m+1)u}{1+(n+m+1)u}\, e(m, u) = \int_0^1 t^n d_t E(u, t), \qquad n = 0, 1, 2, \cdots,$$

$$\gamma_n^{(m)}(u)f(m, u) = \int_0^1 t^n d_t F(u, t), \qquad n = 0, 1, 2, \cdots,$$

which are both regular sequences if $u>0$.

We now use a composition theorem [4, p. 201] for these integrals and get

$$\gamma_n^{(m+1)}(u)f(m+1, u) = \int_0^1 t^n d_t E(u, t)\cdot \int_0^1 t^n d_t F(u, t) = \int_0^1 t^n d_t G(u, t),$$

where

$$G(u, t) = E(u, t) + \int_t^1 F(u, t/v)d_vE(u, v),$$

or

$$f(m + 1, u)B^{(m+1)}(u, t) = E(u, t) + \int_t^1 f(m, u)B^{(m)}(u, t/v)d_vE(u, v).$$

Thus

$$B^{(m+1)}(u, t) = h(u, t) + \int_t^1 B^{(m)}(t, t/v)d_vE^*(u, v)$$

where $E^*(u, v) = [v^{m+1+(1/u)}]/[(m+1)u+1]$, and $h(u, t)$ is a bounded function $|h(u, t)| \leqq H$, $0 \leqq u \leqq 1$, $0 \leqq t \leqq 1$. We therefore have

$$| B^{(m+1)}(u, t) | \leqq H + K^{(m)} \int_0^1 d_vE^*(u, v) \leqq K^{(m+1)}, \qquad u > 0, 0 \leqq t \leqq 1,$$

where $K^{(m+1)}$ is a finite constant. Since the function is identically 0 for $u=0$, the result holds also for $u=0$. This completes the induction and the proof of Theorem 4.1.

II. Gronwall summability

5. **Introduction to Part II.** T. H. Gronwall [6] introduced a very general method of summation, based upon two analytic functions, a *mapping function* $z=f(w)$, and a *weight function* $g(w)=\sum_{p=0}^{\infty} b_pw^p$, $b_p \neq 0$. The nth (f, g)-sum of the series $\sum u_p$ is U_n, where U_n is determined from the power series identity in w

$$\sum_{n=0}^{\infty} u_n[f(w)]^n = [g(w)]^{-1}\sum_{n=0}^{\infty} b_nU_nw^n. \tag{5.1}$$

The function $z=f(w)$ is regular for $|w| \leqq 1$ except possibly at $w=1$, and maps $|w| <1$ one-to-one upon a region D interior to $|z| <1$. Under this mapping, $w=0$ corresponds to $z=0$ and $w=1$ to $z=1$. The inverse function $f^{-1}(z)$ is regular on the boundary of D except possibly at $z=1$, and at this point

$$1 - w = (1 - z)^{\lambda}[a + a_1(1 - z) + \cdots], \qquad \lambda \geqq 1, a >0. \tag{5.2}$$

The function $g(w)$ has the form

$$g(w) = (1 - w)^{-\alpha} + \gamma(w), \qquad \alpha > 0,$$

where $\gamma(w)$ is regular for $|w| \leqq 1$; and $g(w) \neq 0$ for $|w| <1$.

The series $\sum u_n$ is said to be (f, g)-*summable* to s if $\lim_{n\to\infty} U_n = s$.

Special (f, g)-methods are (C, k) for k real and greater than -1; (E, β) (Euler-Knopp) for $0<\beta \leqq 1$; de la Vallée Poussin summability (V); and

a generalized (V)-summability, (Vk), introduced by Gronwall. Recently C. Birindelli [1] has shown that a method of summation of Obrechkoff is the (f, g)-method for which $f=1-(1-w)^{1/2}$, $g=(1-w)^{-1/2}$. We show in §7 that a method of summation introduced by W. A. Mersman [8] is a special (f, g)-method.

Some of the important properties of (f, g)-summability are the following:

(a) If (f, g), (f_1, g_1) are two Gronwall means with map regions D, D_1 and with exponents λ, λ_1 (cf. (5.2)), then $(f, g)\supset(f_1, g_1)$ if $\lambda>\lambda_1$, and D is interior to D_1.

(b) If $\lambda>1$, then $(f, g)\supset(C, k)$, $k>-1$.

(c) The exact domain in which (f, g) sums the geometrical series $\sum x^n$ to the sum $1/(1-x)$ is the interior of the region bounded by the curve

$$x = 1/f(e^{i\theta}), \qquad -\pi \leqq \theta \leqq +\pi.$$

(d) If $\sum u_n$ is (f, g)-summable to s, then $\phi(z)=\sum u_n z^n$ is holomorphic inside the map region D, and $\phi(z)\to s$ uniformly as $z\to 1$ over every path of z interior to D which reaches 1 inside the sector $z=1-re^{i\theta}$, $-\theta_0<\theta<+\theta_0$, $\theta_0<\pi/2\lambda$, $r\geqq 0$, λ defined in (5.2).

The properties (a), (b), (d) were established by Gronwall. Property (c) was established for (Vk) by Gronwall, and extended to the general case by C. Birindelli [2].

The main problem which we consider here is as follows: *to determine all means which are common to the class of Hausdorff means and the class of Gronwall means*. We find these to be the means $[H, c_n]$ where

$$c_n = \beta^n/C_{n+\alpha,n}, \qquad 0 < \beta \leqq 1,\ \alpha \geqq 0, \tag{5.3}$$

the Gronwall mean identical with this Hausdorff mean being (f, g) where

$$f(w) = \frac{\beta w}{1-(1-\beta)w}, \qquad g(w) = (1-w)^{-\alpha-1}. \tag{5.4}$$

6. Hausdorff means which are also Gronwall means. If we suppose that the Hausdorff mean $[H, c_n]$ is the Gronwall mean (f, g), then if we put $u_n=0$ for $n\neq k$, $u_k=1$, in (5.1) we find the relation

$$[f(w)]^k = [g(w)]^{-1}\sum_{n=k}^{\infty} \alpha_{n,k}w^n, \qquad \alpha_{n,k} = \sum_{p=k}^{n} C_{n,p}\Delta^{n-p}c_p, \tag{6.1}$$

which must hold identically in w for $k=0, 1, 2, \cdots$. We shall suppose that $[H, c_n]$ is regular, so that $c_0=1$. Then, by (6.1) with $k=1$, we have

$$f(w) = [g(w)]^{-1}\sum_{n=0}^{\infty} b_n(1-\Delta^n c_0)w^n. \tag{6.2}$$

To determine b_n, put $f(w)=a_1w+a_2w^2+\cdots$ in (6.1). Then we find the rela-

tions $b_n c_n = b_0 a_1^n$, $n=0, 1, 2, \cdots$. If $a_1=0$, then $c_1=c_2=\cdots=0$ inasmuch as b_n must be different from 0, and hence by (6.2) we would have $f(w)\equiv 0$ which is impossible since $f(1)=1$. Consequently, $a_1\neq 0$, and therefore $c_n\neq 0$, $n=1, 2, 3, \cdots$; and $b_n=b_0 a_1^n/c_n$, $n=0, 1, 2, \cdots$. Now it is clear that (5.1) is unaltered if b_n is replaced by kb_n, $n=0, 1, 2, \cdots$, where k is any constant not 0. It follows that there is no restriction in assuming that $b_0=1$. Hence we have the following *necessary* conditions for the (regular) Hausdorff mean $[H, c_n]$ to be the same as the Gronwall mean (f, g):

$$\text{(6.3, i)} \qquad c_n \neq 0, \qquad n = 0, 1, 2, \cdots, \qquad c_0 = 1;$$

$$\text{(6.3, ii)} \qquad g(w) = \sum_{n=0}^{\infty} (\theta w)^n / c_n;$$

$$\text{(6.3, iii)} \qquad f(w) = 1 - [g(w)]^{-1} \sum_{n=0}^{\infty} \Delta^n c_0 (\theta w)^n / c_n,$$

where θ is a parameter unequal to 0 so chosen that $f(1)=1$.

To obtain other *necessary* conditions, divide both members of (5.1) by $1-f(w)$, put $s_n=u_0+u_1+\cdots+u_n=0$, $n\neq k$, $s_k=1$, $U_n=\sum_{p=0}^{n} C_{n,p}\Delta^{n-p}c_p\cdot s_p$, $\theta w=t$, $f(w)=F(t)$, and we obtain the equation

$$\text{(6.4)} \qquad (F(t)/t)^k = \left(\sum_{n=0}^{\infty} c_n^{-1} \Delta^n c_0 t^n\right)^{-1} \sum_{n=0}^{\infty} c_{n+k}^{-1} C_{n+k,k} \Delta^n c_k t^n.$$

This obviously holds when $k=0$. If it is to hold for all k it must hold for $k+1$ when it holds for k. Hence, on multiplying both members of (6.4) by $F(t)/t$ we find that a necessary and sufficient condition for (6.4) to hold for all k is that

$$\text{(6.5)} \qquad \begin{aligned} &\sum_{n=0}^{\infty} c_{n+1}^{-1}(1-\Delta^{n+1}c_0)t^n \cdot \sum_{n=0}^{\infty} c_{n+k}^{-1} C_{n+k,k}\Delta^n c_k t^n \\ &\qquad \equiv \sum_{n=0}^{\infty} c_n^{-1} t^n \cdot \sum_{n=0}^{\infty} c_{n+k+1}^{-1} C_{n+k+1,k+1}\Delta^n c_{k+1} t^n, \qquad k = 0, 1, 2, \cdots. \end{aligned}$$

On equating the coefficients of the first power of t on either side we obtain the relation

$$c_{k+1}^{-1}(k+1)(c_k - c_{k+1}) + c_2^{-1}(2c_1 - c_2) = c_{k+2}^{-1}(k+2)(c_{k+1} - c_{k+2}) + c_1^{-1}.$$

This serves as a recursion relation to determine c_k parametrically, whence we find that c_k must have the form $\beta^k/C_{k+\alpha,k}$.

When this value of c_k is substituted in (ii), (iii) of (6.3) we get: $F(t)=\beta t/[\beta-(1-\beta)t]=f(t/\theta)$, $f(w)=\beta w\theta/[\beta-(1-\beta)w\theta]$. If we take $\theta=\beta$ we find for $f(w)$, $g(w)$ the values (5.4). If these functions are to satisfy the conditions of Gronwall we must have, first of all, α real and $\alpha+1>0$. But since

$[H, c_n]$ is a regular Hausdorff mean, $\Re(\alpha) \geqq 0$. Hence we must have α real and greater than or equal to 0.

We now determine β so that the map of $|w| < 1$ by $z = f(w)$ shall lie in $|z| < 1$. Put $w = e^{i\theta}$, $\beta = p + iq$, and we see that $|z| \leqq 1$ if and only if $p + [q/\tan(\theta/2)] \leqq 1$; and hence $q = 0, \beta = p \leqq 1$. The map of $|w| = 1$ is a circle in the z-plane whose intercepts on the real axis are $z = -\beta/(2-\beta)$, $z = 1$. In order that this map should contain the origin it is necessary that $\beta \geqq 0$. Since $\beta = 0$ is clearly excluded, we therefore have $0 < \beta \leqq 1$.

We have shown that c_n must have the form (5.3). Consequently

$$c_n = \int_0^1 u^n d\phi(u), \qquad n = 0, 1, 2, \cdots, \tag{6.6}$$

where

$$\phi(u) = \begin{cases} 1 - (1 - u\beta^{-1})^\alpha, & u \leqq \beta, \\ 1, & u > \beta. \end{cases}$$

Thus $[H, c_n]$ is a regular Hausdorff mean.

It remains to be proved that when c_n is given by (6.6) then (6.4) holds for all values of k. On making use of the integral representation (6.6) we find that (6.4) will hold if

$$\left[\frac{1 - w}{1 - (1 - \beta)w}\right]^{k+1} = \alpha C_{k+\alpha,k} \int_0^1 \frac{u^k(1 - u)^{\alpha-1} du}{[1 + \beta w(1 - w)^{-1} u]^{k+\alpha+1}}.$$

But the right member is equal to

$$\alpha C_{k+\alpha,k} \frac{\Gamma(k + 1)\Gamma(\alpha)}{\Gamma(k + \alpha + 1)} \sum_{n=0}^{\infty} C_{n+k,n}[- \beta w/(1 - w)]^n = [1 + \beta w(1 - w)^{-1}]^{-k-1},$$

which is equal to the left member. It now follows that $[H, c_n]$ is the Gronwall mean (f, g), where f, g are given by (5.4).

We have proved the following theorem.

THEOREM 6.1. *A necessary and sufficient condition in order for a regular Hausdorff mean $[H, c_n]$ to be a Gronwall mean (f, g) is that*

$$c_n = \beta^n/C_{n+\alpha,n} = \int_0^1 u^n d\phi(u), \qquad n = 0, 1, 2, \cdots, \alpha \geqq 0, 0 < \beta \leqq 1,$$

where $\phi(u) = 1 - (1 - u\beta^{-1})^\alpha$ if $u \leqq \beta$, $\phi(u) = 1$ if $u > \beta$. The Gronwall mean which is the same as this has f, g given by (5.4).

We note that the domain in which this mean sums the geometrical series, and which is given by (c) of §5, is the circular region $|z + (1-\beta)/\beta| < 1/\beta$. This result furnishes additional evidence in support of the conjecture [5, p.

205] that a necessary and sufficient condition that a Hausdorff mean sum a power series outside its circle of convergence is that the mass function be constant in the neighborhood of 1.

7. **Mersman summability**. In a recent paper, W. A. Mersman [8] studied the transformation

$$(7.1)\qquad U_n = (\tfrac{1}{2})^{2n}\sum_{p=0}^{n} C_{2n+k,n-p}\cdot(u_0+u_1+\cdots+u_p),\qquad n=0,1,2,\cdots.$$

He proved that this defines a regular method of summation, which we shall call (M)-summability. He found that $(M)\supset(C,k)$ for $k=1, 2$; (M) includes an Euler-Knopp method; and he determined the domain in which (M) sums the geometrical series. We shall now prove the following theorem.

THEOREM 7.1. *(M)-summability is (f, g)-summability with*

$$(7.2)\qquad f(w)=\frac{1-(1-w)^{1/2}}{1+(1-w)^{1/2}},\qquad g(w)=(1-w)^{-1}.$$

Proof. It is required to show that when these values of $f(w)$, $g(w)$, and U_n from (7.1) are substituted in (5.1), the latter becomes a power series identity. This can be done in exactly the same way that Gronwall established the corresponding theorem for de la Vallée Poussin summability (V) determined by

$$(7.3)\qquad U_n=\sum_{p=0}^{n}\frac{(n!)^2}{(n-p)!(n+p)!}u_p,\qquad n=0,1,2,\cdots,$$

with $f(w)$, $g(w)$ given by

$$(7.4)\qquad f(w)=\frac{1-(1-w)^{1/2}}{1+(1-w)^{1/2}},\qquad g(w)=(1-w)^{-1/2}.$$

On making the indicated substitutions, and putting $u_n=0$, $n\neq k$, $u_k=1$, we find that the following identity must be verified

$$(7.5)\qquad (1-w)^{-1}[f(w)]^k=\sum_{n=k}^{\infty}(\tfrac{1}{2})^{2n}\sum_{p=k}^{n}C_{2n+1,n-p}w^n,\qquad k=0,1,2,\cdots.$$

One may show directly that this holds for $k=0, 1$. Then by means of the identity

$$[f(w)]^{k+1}=[f(w)]^k(4w^{-1}-2)-[f(w)]^{k-1},\qquad k=1,2,3,\cdots,$$

one may prove (7.5) by mathematical induction for all values of k. On substituting (7.5) in (5.1) the proof may be completed exactly as in Gronwall's proof.

On comparing (7.2) and (7.4) we see that the map function $f(w)$ is the same for (M)- and (V)-summabilities, while $g(w)$ is of the form $(1-w)^{-\alpha}$ for

both methods, with α having a larger value in the case of (M)-summability. It therefore follows from a theorem of Birindelli [2] that $(V)\subset(M)$, $(M)\not\subset(V)$ This will also follow from results to be given in Part III.

III. A METHOD OF SUMMATION ARISING FROM AN ALGORITHM OF SCHUR

8. **Introduction to Part** III. The starting point of Part III is the following theorem of Wall [13].

THEOREM 8.1. *Let $e(x)$ be any function analytic and with modulus less than or equal to 1 for $|x|<1$, which is real when x is real. Then there exists a function $F(z)$ of the form*

$$F(z) = \int_0^1 \frac{d\phi(u)}{1+zu}, \tag{8.1}$$

$\phi(u)$ bounded, monotone non-decreasing, $0\leqq u\leqq 1$, with modulus less than or equal to 1 for $|z|<1$, such that

$$\tfrac{1}{2}(1-x)\frac{1-e(x)}{1+xe(x)} = F(z), \quad z = 4x/(1-x)^2, \quad |x| < 1. \tag{8.2}$$

Conversely, if $F(z)$ is any function of the form (8.1) *with modulus less than or equal to 1 for $|z|<1$, then there exists a function $e(x)$ of the kind described above such that* (8.2) *holds.*

The theorem grows out of an algorithm used by Schur [10] in his work on functions bounded in the unit circle. This algorithm yields a continued fraction representation for the function in the left member of (8.2), and the continued fraction is in turn equal to a function of the form (8.1) with modulus less than or equal to 1 for $|z|<1$. (For details, and a discussion of some consequences of this theorem, see the paper of Wall [13].)

We shall denote by E the class of functions $e(x)$ described in this theorem.

Put $[1-e(x)]/[1+xe(x)]=\sum c_n(-x)^n$, $F(z)=\sum c_n(-z)^n$ in (8.2). Considering the result as a power series identity in x, and equating coefficients of like powers of x on either side, we obtain a transformation of the form

$$C_n = \sum_{p=0}^{n} T_{n,p}c_p, \qquad n = 0, 1, 2, \cdots. \tag{8.3}$$

From the above theorem it follows that if $e(x)\in E$, then the transformation (8.3) *carries the sequence $\{c_p\}$ into a totally monotone sequence $\{C_n\}$*. In §9 we have formulated this in such a way as to characterize a class of functions having positive real part for $|x|<1$.

We find, moreover, that if $\{c_n\}$ is any moment sequence, then $\{C_n\}$ is also a moment sequence, so that $[H, C_n\}$ is a Hausdorff mean. The set of all these means forms a regular manifold in the sense of Part I.

Considered as a transformation on the series $\sum c_n$ to the series $\sum C_n$, (8.3) defines a regular method of summation, which we call *Schur summability* and denote by (S). It turns out that $(S)\approx(M)$, where (M) is the method of Mersman discussed in §7. Thus (S) is equivalent to a Gronwall method.

9. **Schur summability**. If we put $f(x)=\sum c_n(-x)^n$, $F(z)=\sum C_n(-z)^n$ in the relation

$$(1/2)(1-x)f(x)=F(z),\qquad z=4x/(1-x)^2, \tag{9.1}$$

write both members as power series in x, and then equate coefficients of like powers of x, we obtain the relation

$$(-1)^n c_n=\sum_{p=0}^{n}(-1)^p(1/2)^{2p+1}C_{n+p,2p}C_p,\qquad n=0,1,2,\cdots. \tag{9.2}$$

If $(-1)^n c_n=a_n$, $(\frac{1}{2})^{2p+1}(-1)^pC_p=A_p$, this gives

$$(-1)^kA_k=t_{k,0}a_0-t_{k,1}a_1+\cdots+(-1)^kt_{k,k}a_k,\qquad k=0,1,2,\cdots,$$

$t_{k,p}$ being a certain determinant, namely,

$$t_{k,p}=J_{2p,k-p},\qquad p=0,1,2,\cdots,k,$$

where

$$J_{k,0}=1,\ J_{k,m}=\begin{vmatrix} C_{k+1,1}, & C_{k+2,0}, & 0, & 0, & \cdots, & 0\\ C_{k+2,2}, & C_{k+3,1}, & C_{k+4,0}, & 0, & \cdots, & 0\\ & & & \cdot & & \cdot\\ & & & \cdot & & \cdot\\ \cdot & \cdot & \cdot & \cdot & \cdots & 0\\ \cdot & \cdot & \cdot & \cdot & \cdots & C_{k+2m-2,0}\\ C_{k+m,m}, & C_{k+m+1,m-1}, & \cdot & \cdot & \cdots, & C_{k+2m-1,1} \end{vmatrix},$$

$k=0,1,2,\cdots$, $m=1,2,3,\cdots$. On subtracting the $(m-k)$th row from the $(m-k+1)$th row in succession for $k=1,2,3,\cdots,m-1$, we readily obtain

$$J_{k,m}=J_{k-1,m}+J_{k+1,m-1}=\sum_{p=1}^{k+1}J_{p,m-1}. \tag{9.3}$$

This holds for $k\geqq 1$, $m=1,2,3,\cdots$. Hence it follows that

$$J_{k,m}=(k+1)(k+m+1)^{-1}C_{k+2m,m},\qquad k\geqq 1,\ m=0,1,2,\cdots. \tag{9.4}$$

This holds also for $k=0$. In fact, if $f(x)=1$ in (9.1), i.e., $c_0=1$, $c_n=0$, $n>0$, then

$$F(z)=z^{-1}\left[(1+z)^{1/2}-1\right]=\sum_{p=0}^{\infty}(\tfrac{1}{2})^{2p+1}(p+1)^{-1}C_{2p,p}(-z)^p,$$

and consequently $t_{n,0}=J_{0,n}$ is given by (9.4) with $k=0$. We may now write down at once the inverse of the transformation (9.2):

$$(9.5)\qquad C_n = \sum_{p=0}^{n} T_{n,p}c_p, \quad T_{n,p} = (\tfrac{1}{2})^{2n+1}(2p+1)(n+p+1)^{-1}C_{2n,n-p}.$$

It is desirable to express this as a transformation on the partial sums $S_n = C_0 + C_1 + \cdots + C_n$, $s_n = c_1 + c_0 + \cdots + c_n$ of the series $\sum C_n$, $\sum c_n$. To do this, let $S_n = \sum_{p=0}^{n} b_{n,p}s_p$, $n = 0, 1, 2, \cdots$, and it will be seen that

$$(9.6)\qquad b_{n,p} = b_{n-1,p} + T_{n,p} - T_{n,p+1},$$

$p = 0, 1, 2, \cdots, n$, $b_{n-1,n} = T_{n,n+1} = 0$. One may then show by mathematical induction that $b_{n,p} = (\frac{1}{2})^{2n+1}C_{2n+2,n-p}$. We now make the following definition.

Definition. *The transformation*

$$(9.7)\qquad S_n = \sum_{p=0}^{n} b_{n,p}\, s_p, \quad b_{n,p} = (\tfrac{1}{2})^{2n+1}C_{2n+2,n-p}, \quad n = 0, 1, 2, \cdots,$$

is called the (*S*)*-transformation; and the method of summation which assigns to the sequence* $\{s_n\}$ *the value* $\lim_{n\to\infty} S_n$ *when the latter limit exists, is called* (*S*)*-summability.*

From Theorem 8.1 and from the way in which (*S*)-summability was defined we have this theorem:

Theorem 9.1. *Let E denote the class of functions* $e(x)$ *which are analytic and have moduli less than or equal to* 1 *for* $|x| < 1$, *and which are real when x is real. Then* $e(x)$ *is in E if and only if* (*S*) *transforms the sequence* $s_n = c_0 + c_1 + \cdots + c_n$ *defined by* $[1 - e(x)]/[1 + xe(x)] = \sum c_n(-x)^n$ *into a sequence* $S_n = C_0 + C_1 + \cdots + C_n$, *convergent and with limit*(9) *less than or equal to* 1, *and such that* $\{C_n\}$ *is a totally monotone sequence.*

A theorem equivalent to Theorem 9.1 is as follows:

Theorem 9.2. *Let K denote the class of power series, with real coefficients, of the form*

$$(9.8)\qquad k(x) = \tfrac{1}{2} - D_0x + D_1x^2 - D_2x^3 + \cdots,$$

convergent and having real part greater than or equal to 0 *for* $|x| < 1$. *Then* $k(x)$ *is in K if and only if* (*S*) *transforms the sequence* $\{1 - D_n\}$ *into a sequence* $S_n = C_0 + C_1 + \cdots + C_n$ *convergent and with limit less than or equal to* 1, *and such that* $\{C_n\}$ *is a totally monotone sequence.*

Proof. Let $e(x)$, $k(x)$ be two power series with real coefficients related by the equation $k(x) = 1/2 - xe(x)[1 + xe(x)]^{-1}$. Then it is easily seen that $k(x)$ is in K if and only if $e(x)$ is in E. Moreover, if $[1 - e(x)]/[1 + xe(x)] = \sum c_n(-x)^n$, and $k(x)$ is of the form (9.8), then $1 - D_n = c_0 + c_1 + \cdots + c_n$, $n = 0, 1, 2, \cdots$. The theorem now follows at once from Theorem 9.1.

(9) The modulus of $F(z)$ is less than or equal to 1 for $|z| < 1$ if and only if $\sum C_n \leq 1$ [11, p. 181].

10. **The (S)-transformation and moment sequences.** By an application of the work in Part I we shall prove the following theorem.

THEOREM 10.1. *The (S)-transformation in the form (9.5) transforms any moment sequence $\{c_n\}$ into another moment sequence $\{C_n\}$ such that $[H, C_n]$ is an essentially regular Hausdorff mean.*

Proof. Let $c_p=\int_0^1 u^p d\phi(u)$, $\phi(u)\in BV[0, 1]$, $p=0, 1, 2, \cdots$. Then by (9.5) we see that

$$C_n = \int_0^1 \beta_n(u)d\phi(u), \qquad n = 0, 1, 2, \cdots,$$

where $\beta_n(u)=\sum_{p=0}^n T_{n,p}u^p$. To prove the theorem it is therefore required to show that the sequence $\{\beta_n(u)\}$ is the basis of a regular manifold.

Let $0\leqq u\leqq 1$. Then $e(x)=u=$constant is a function in the class E of Theorem 9.1. Also,

$$[1 - e(x)]/[1 + xe(x)] = (1 - u)/(1 + xu) = (1 - u)\sum_{p=0}^{\infty} u^p(- x)^p.$$

Hence it follows from Theorem 9.1 that the sequence $\{(1-u)\beta_n(u)\}$ is totally monotone, and consequently $\{\beta_n(u)\}$ is totally monotone if $0\leqq u<1$. It is also totally monotone for $u=1$ by reason of continuity. Therefore $\{\beta_n(u)\}$ is a totally monotone basis of a manifold $M[\beta_n(u)]$.

To show that $M[\beta_n(u)]$ is regular, we put $f(x)=1/(1+ux)$ in (9.1) and find that the corresponding value of $F(z)$ is

$$(10.1) \qquad [(1 - u) + (1 + u)(1 + z)^{1/2}]^{-1} = \sum_{p=0}^{\infty} \beta_p(u)(- z)^p, \quad 0 \leqq u \leqq 1.$$

We may write this in the form

$$(10.2) \qquad [(1 - u) + (1 + u)(1 + z)^{1/2}]^{-1} = \int_0^1 \frac{d_tB(u, t)}{1 + zt},$$

where $B(u, t)$ is a monotone function of t, $0\leqq t\leqq 1$, for each u, $0\leqq u\leqq 1$, determined by the equations $\beta_p(u)=\int_0^1 t^p d_tB(u, t)$, $p=0, 1, 2, \cdots$. We see by inspection that (10.2) tends to 0 as z tends to ∞ through positive real values, in consequence of which $B(u, t)$ is continuous at $t=0$. This in turn implies that $\lim_{n\to\infty} \Delta^n\beta_0(u)=0$, $0\leqq u\leqq 1$. Therefore, by Theorem 2.2, $M[\beta_n(u)]$ is regular and the proof of the theorem is complete.

With a view toward finding out other properties of the means in the manifold $M[\beta_n(u)]$ we obtained $B(u, t)$ explicitly from (10.2), using the Stieltjes [9, p. 372] inversion formula. The result is

$$B(u, t) = \frac{1}{\pi}\int_0^t \frac{1 + u}{(1 + u)^2 - 4us}\left(\frac{1}{s} - 1\right)^{1/2} ds, \qquad 0 \leqq u \leqq 1,\ 0 \leqq t \leqq 1.$$

By means of a known [4, p. 202] necessary and sufficient condition for a Hausdorff mean to include (C, n) for integral n, we find that special means in $M[\beta_n(u)]$ include $(C, 1)$ but not $(C, 2)$. That this is not always the case can be seen from the fact that the regular mean $[H, 2\beta_n(1)]$, which is in the manifold, does not include $(C, 1)$. In fact, $\beta_n(1):(n+1)^{-1} \geqq (n+1)/4n^{1/2}$, and consequently this ratio cannot be the nth member of a moment sequence. Hence $(C, 1) \not\subset [H, 2\beta_n(1)]$.

11. **The relation of (S)-summability to Gronwall summability.** On modifying the notation to facilitate comparison with the work in Part II we find that (S)-summability is determined by the relation

$$\sum_{n=0}^{\infty} u_n[f(w)]^n = \frac{1 + (1 - w)^{1/2}}{(1 - w)^{-1}} \sum_{n=0}^{\infty} U_n w^n, \quad f(w) = \frac{1 - (1 - w)^{1/2}}{1 + (1 - w)^{1/2}}, \tag{11.1}$$

in the same way that (f, g)-summability is determined by (5.1). This is therefore not a Gronwall method inasmuch as the function $1/(1-w)[1+(1-w)^{1/2}]$ does not satisfy the conditions required by Gronwall for $g(w)$. *However, (S) is equivalent to (M), which we showed to be an (f, g)-method.* In fact, if we put $[1+(1-w)^{1/2}]\sum_{n=0}^{\infty} U_n w^n = \sum_{n=0}^{\infty} U_n^* w^n$, then (11.1) becomes

$$\sum_{n=0}^{\infty} u_n[f(w)]^n = \frac{1}{(1 - w)^{-1}} \sum_{n=0}^{\infty} U_n^* w^n,$$

which is formally the same as (M)-summability. Now it is easily seen that $\lim_{n\to\infty} U_n = s$ if and only if $\lim_{n\to\infty} U_n^* = s$, and consequently $(S) \approx (M)$. In fact, if $1+(1-w)^{1/2} = \sum v_n w^n$, then $U_n^* = v_0 U_n + v_1 U_{n-1} + \cdots + v_n U_0$, and since $\sum v_n$ converges absolutely and $\sum v_n = 1$, it follows that lim $U_n = s$ implies lim $U_n^* = s$, so that $(S) \subset (M)$. Since $\sum U_n w^n = w^{-1}[1-(1-w)^{1/2}]\sum U_n^* w^n$, it follows similarly that $(M) \subset (S)$.

Let $S_n(s_0, s_1, \cdots)$ denote the nth (S)-sum of the sequence $\{s_n\}$, and $M_n(s_0, s_1, \cdots)$ the nth (M)-sum of the same sequence. Then from (9.7), (7.1) it follows that $S_n(s_0, s_1, \cdots) = (1/2)M_n(s_0, s_1, \cdots) + (1/2)M_n(0, s_0, s_1, \cdots)$. From this and from the fact that $(S) \approx (M)$ we have this theorem:

THEOREM 11.1. *If the (S)-limit of the sequence $s_0, s_1, s_2, \cdots$ is s, then the (S)-limit of the sequence $0, s_0, s_1, \cdots$ is also s.*

A number of properties of (S)-summability follow from the general theorems about Gronwall summability: e.g., $(V) \subset (S)$, $(S) \not\subset (V)$; (S) sums the geometrical series in the same domain as that in which (V) sums this series; $(S) \supset (C, k)$ for $k > -1$. In the next paragraph we shall prove some of these things by direct methods.

12. **Some properties of (S)-summability.** *Using the relationships obtained in §§9–10 we shall prove the following propositions*: (a) (S) *is regular*, (b) (S) *sums the geometrical series* $\sum x^n$ *to* $1/(1-x)$ *inside the curve*

$$(12.1)\qquad r = 2 - \cos\theta + [(1 - \cos\theta)(3 - \cos\theta)]^{1/2},$$

while (S) *does not sum the geometrical series outside or upon this curve*; (c) $(V)\subset(S)$; (d) $(S)\not\subset(V)$.

Proof of (a). It is required to show that the numbers $b_{n,p}$ of (9.7) satisfy the *regularity conditions*:

$$(12.2,\text{ i})\qquad \sum_{p=0}^{n} | b_{n,p} | < M \textit{ for every } n,\ M \textit{ independent of } n;$$

$$(12.2,\text{ ii})\qquad \sum_{p=0}^{n} b_{n,p} \textit{ tends to } 1 \textit{ as } n \textit{ tends to } \infty;$$

$$(12.2,\text{ iii})\qquad \lim_{n=\infty} b_{n,p} = 0 \textit{ for every } p.$$

From (9.6) we have $\sum_{p=0}^{n} b_{n,p} = \sum_{p=0}^{n} T_{p,0}$. As n tends to ∞ this tends to 1 inasmuch as the power series $[(1+z)^{1/2}-1]/z = \sum_{p=0}^{\infty} T_{n,p}(-z)^p$ converges and has the sum 1 when $z=-1$. This proves (ii). Since the $b_{n,p}$'s are all positive, (i) holds by virtue of (ii). To prove (iii), we use the relation

$$(12.3)\qquad b_{n,p} = \sum_{k=p}^{n} T_{k,p} - \sum_{k=p+1}^{n} T_{k,p+1},$$

which is a consequence of (9.6). As $n\to\infty$, the two sums of positive terms on the right have finite limits. For if $f(x) = (-x)^p$ in (9.1) there results the equation

$$(12.4)\qquad \left(\frac{(1 + z)^{1/2} - 1}{z}\right)^{2p+1} = \sum_{k=p}^{\infty} T_{k,p}(- z)^{k-p};$$

and this series converges for $z = -1$, being the Cauchy product of $2p+1$ absolutely convergent series. Since the value of the function on the left is 1 when $z = -1$ it follows that $\sum_{k=p}^{\infty} T_{k,p} = 1$, $p = 1, 2, 3, \cdots$. Therefore, by (12.3) we see that the last of the regularity conditions is satisfied, and therefore (S) is regular.

We note for future reference that

$$(12.5)\qquad \sum_{k=p}^{\infty} T_{k,p} z^{k-p} = (1/2)^{2p+1} F(p + 1/2,\ p + 1,\ 2p + 2;\ z),$$

where $F(\alpha, \beta, \gamma; z)$ is the hypergeometric series.

Proof of (b). On replacing c_p by x^p in (9.5) we find that in order to show that the geometrical series $\sum x^p$ is (S)-limitable for a particular value of x, it is required to prove that the series $\sum \beta_p(x)$ is convergent, where $\beta_n(x) = \sum_{p=0}^{n} T_{n,p} x^p$. Using the recursion formula (9.3) one may show that these polynomials satisfy the relation.

(12.6) $\beta_p(x) = w[\beta_{p-1}(x) - (1+x)^{-1}T_{p-1}(0)],\ w = (1+x)^2/4x,\ \beta_0(x) = 1/2,$

and consequently

$$\beta_n(x) = (1/2)w^n - w(1+x)^{-1}(T_{0,0}w^{n-1} + T_{1,0}w^{n-2} + \cdots + T_{n-2,0}w + T_{n-1,0}).$$

Now the last quantity in parentheses is the coefficient of t^{n-1} in the power series in t for the function $[1-(1-t)^{1/2}]/t(1-wt)$, which is convergent for $t=1$ provided $|w|<1$. It follows that if $|w|<1$,

$$\sum \beta_n(x) = (1/2)(1-w)^{-1} - w(1-w)^{-1}(1+x)^{-1} = 1/(1-x).$$

On the other hand, if $|w| \geqq 1$, $|1+x| \geqq 2$, we have

$$\beta_n(x) = w^n[1/2 - (1+x)^{-1}(T_{0,0} + T_{1,0}w^{-1} + \cdots + T_{n-1,0}w^{1-n})].$$

Inasmuch as the series $\sum_{p=0}^{\infty} T_{p,0}w^{-p}$ converges and has a sum numerically less than or equal to 1 it follows that

$$|\beta_n(x)| \geqq 1/2 - |(1+x)^{-1}|, \qquad |w| \geqq 1, \qquad |1+x| \geqq 2.$$

Now $\sum \beta_n(1)$ evidently diverges, being the series for $(1/2)(1+z)^{-1/2}$ evaluated at $z=-1$. In any other case where $|w| \geqq 1$, $|1+x| \geqq 2$ we have $|\beta_n(x)| \geqq d > 0$, where d is a constant independent of n, and hence $\sum \beta_n(x)$ diverges. Since the curve $|w| = 1$, $|1+x| \geqq 2$ is given in polar form by (12.1), and since the interior of this curve corresponds to $|w|<1$ and the exterior to $|w|>1$, the proof of (b) is now complete.

Proof of (c). The nth (V)-sum of the sequence $\{s_n\}$ is given by

$$V_n = \sum_{p=0}^{n} \frac{(n!)^2(2p+1)}{(n-p)!(n+p+1)!} s_p. \tag{12.7}$$

On eliminating $s_0, s_1, \cdots, s_n$ between (12.7) and (9.7) we obtain a relation of the form $S_n = \sum_{p=0}^{n} e_{n,p} V_p$, where

$$(\tfrac{1}{2})^{2n+1} C_{2n+2,n-p} = \sum_{k=p}^{n} e_{n,k} \frac{(k!)^2(2p+1)}{(k-p)!(k+p+1)!}.$$

We shall prove that

$$e_{n,p} = (\tfrac{1}{2})^{2n+1} \frac{(2p)!}{(p!)^2} C_{2n-2p+1,n-p}. \tag{12.8}$$

To do this it will suffice to verify the identity

$$C_{2n+2,n-k} = (2k+1) \sum_{p=k}^{n} (2p+1)^{-1} C_{2p+1,p-k} C_{2n-2p+1,n-p}.$$

The quantity on the right may be identified as the coefficient of x^{n-k} in the

product $F(3/2, 1, 2; 4x)F(k+1/2, k+1, 2k+2; 4x)$, while the left member is the coefficient of x^{n-k} in $F(k+3/2, k+2, 2k+3; 4x)$. Thus we must establish the power series identity

$$F(k + 3/2, k + 2, 2k + 3; z) = F(3/2, 1, 2; z)\cdot F(k + 1/2, k + 1, 2k + 2; z).$$

Since $D_zF(k+1/2, k+1, 2k+2; z)=(1/4)(2k+1)F(k+3/2, k+2, 2k+3; z)$, the relation to be established reduces to

$$F(k + 1/2, k + 1, 2k + 2; z) = \left(2\cdot\frac{1 - (1 - z)^{1/2}}{z}\right)^{2k+1},$$

which is an identity by (12.4), (12.5).

In order to show that $(V)\subset(S)$, we must show that the regularity conditions (12.2) are satisfied by the numbers $e_{n,p}$. Condition (iii) is clearly satisfied; and (i) will follow from (ii). We have

$$\sum_{p=0}^{n} e_{n,p} = (\tfrac{1}{2})^{2n+1}\sum_{p=0}^{n} C_{2n-2p+1,n-p}C_{2p,p}.$$

This is seen to be the coefficient of x^n2^{-2n-1} in the power series in x for the function $2[1-(1-4x)^{1/2}]/4x(1-4x)$. Hence it is required to show that the coefficient of x^n in the power series in x for the function $(1-x)^{-1}[1-(1-x)^{1/2}]/x$ has the limit 1 for $n=\infty$. But if this coefficient is denoted by d_n, we have

$$d_n = v_0q_n + v_1q_{n-1} + \cdots + v_nq_0, \tag{12.9}$$

where $q_0=q_1=\cdots=1$, and $[1-(1-x)^{1/2}]/x=\sum v_nx^n$. Since $\sum v_n$ converges absolutely and $\sum v_n=1$, it follows that (12.9) constitutes a regular transformation of the sequence $\{q_n\}$ into the sequence $\{d_n\}$. Hence, inasmuch as $q_n=1$ it follows that $\lim d_n=1$, as was to be proved. We have completed the proof that $(V)\subset(S)$.

Proof of (d). To show that $(S)\not\subset(V)$ we shall show that there is at least one sequence which is summable (S) but which is not summable (V). For that purpose it will suffice to show that if we put $V_p=(-1)^p$ in the relation $S_n=\sum_{p=0}^{n}e_{n,p}V_p$ of the preceding proof, then $\lim S_n$ exists and is finite. We find that

$$4^{-1}(1 - t^2)^{-1/2}[(1 - t)^{1/2} + 1] = S_0 + S_1t + S_2t^2 + \cdots,$$

where $S_n=\sum_{p=0}^{n}(-1)^pe_{n,p}$. Thus if $(1-t)^{1/2}+1=\sum v_pt^p$, $4^{-1}(1-t^2)^{-1/2}=\sum q_pt^p$, then $S_n=v_0q_n+v_1q_{n-1}+\cdots+v_nq_0$. This is a regular transformation of the sequence $\{q_n\}$ into the sequence $\{S_n\}$. Hence, inasmuch as $\lim q_n=0$, it follows that $\lim S_n=0$. This establishes (d).

Bibliography

1. Carlo Birindelli, *Sui metodi di Gronwall per la sommazione delle serie*, Annali delle Scuòla Normale Superiore di Pisa (2), vol. 8 (1939), pp. 241–270.

2. ———, *Contributo all'analisi dei metodi di sommazione di Gronwall*, Rendiconti del Circolo Matematico di Palermo, vol. 61 (1937), pp. 157–176.

3. H. E. Bray, *Elementary properties of Stieltjes integrals*, Annals of Mathematics, (2), vol 20 (1918), pp. 177–186.

4. H. L. Garabedian, Einar Hille, and H. S. Wall, *Formulations of the Hausdorff inclusion problem*, Duke Mathematical Journal, vol. 8 (1941), pp. 193–213.

5. H. L. Garabedian and H. S. Wall, *Hausdorff methods of summation and continued fractions*, these Transactions, vol. 48 (1940), pp. 185–207.

6. T. H. Gronwall, *Summation of series and conformal mapping*, Annals of Mathematics, (2), vol. 33 (1932), pp. 101–117.

7. F. Hausdorff, *Summationsmethoden und Momentfolgen*. I and II, Mathematische Zeitschrift, vol. 9 (1921), pp. 74–109 and 280–299.

8. W. A. Mersman, *A new method of summation of divergent series*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 667–673.

9. O. Perron, *Die Lehre von den Kettenbrüchen*, 2d edition, Leipzig and Berlin, 1929.

10. J. Schur, *Ueber Potenzreihen die im Innern des Einheitskreises beschränkt sind*, Journal für die reine und angewandte Mathematik, vol. 147 (1916), pp. 205–232 and vol. 148 (1917), pp. 122–145.

11. H. S. Wall, *Continued fractions and totally monotone sequences*, these Transactions, vol. 48 (1940), pp. 165–184.

12. ———, *A class of functions bounded in the unit circle*, Duke Mathematical Journal, vol. 7 (1940), pp. 146–153.

13. ———, *Some recent developments in the theory of continued fractions*, Bulletin of the American Mathematical Society, vol. 47 (1941), pp. 405–423.

Northwestern University,
Evanston, Ill.

ON CONFORMAL MAPPING OF INFINITE STRIPS

BY

S. E. WARSCHAWSKI

INTRODUCTION

Let S be the strip in the plane of the complex variable $w=u+iv$ defined by the relations

$$\phi_-(u) < v < \phi_+(u), \qquad -\infty < u < +\infty,$$

where $\phi_-(u)$, $\phi_+(u)$ are continuous for $-\infty < u < +\infty$. Let $\theta(u) \equiv \phi_+(u) - \phi_-(u)$ and $\psi(u) \equiv \frac{1}{2}[\phi_+(u) + \phi_-(u)]$. S can be mapped conformally onto the strip $|y| < \pi/2$ of the z-plane, $z=x+iy$, by means of an analytic function $z=Z(w)=X(w)+iY(w)$ in such a manner that $\lim_{u\to+\infty} X(w)=+\infty$. The principal object of this paper is to obtain asymptotic expressions for $Z(w)$ and its derivative $Z'(w)$ as $u\to+\infty$. For this purpose two inequalities concerning the difference $X(w_2)-X(w_1)$ ($w_1=u_1+iv_1$, $w_2=u_2+iv_2$ in S) are established which are similar to certain inequalities of Ahlfors(1), but which, due to some assumptions regarding the smoothness of the boundary of S, yield sharper estimates for large values of u_1 and u_2.

We say that S is an *L-strip*(2) *with the boundary inclination* γ *at* $u=+\infty$, $|\gamma| < \pi/2$, if, for $u_2 > u_1$,

$$\frac{\phi_+(u_2) - \phi_+(u_1)}{u_2 - u_1}, \qquad \frac{\phi_-(u_2) - \phi_-(u_1)}{u_2 - u_1}$$

approach the same limit, $\tan\gamma$, as u_1 and $u_2 \to +\infty$ simultaneously. The two inequalities in question (the "basic inequalities") are then as follows:

I. *If S is an L-strip with the boundary inclination $\gamma=0$ at $u=+\infty$, then*

$$X(w_2) - X(w_1) \leqq \pi \int_{u_1}^{u_2} \frac{1+\psi'^2(u)}{\theta(u)}\,du + \frac{\pi}{12}\int_{u_1}^{u_2} \frac{\theta'^2(u)}{\theta(u)}\,du + o(1),$$

where $o(1)\to 0$ as $u_1, u_2 \to +\infty$, uniformly with respect to v_1 and v_2.

II. *If S is an L-strip as in* I *and if, in addition, $\phi_+'(u)$ and $\phi_-'(u)$ are continuous and of bounded variation for $u_0 \leqq u \leqq +\infty$, then*

$$X(w_2) - X(w_1) \geqq \pi \int_{u_1}^{u_2} \frac{1+\psi'^2(u)}{\theta(u)}\,du - \frac{\pi}{4}\int_{u_1}^{u_2} \frac{\theta'^2(u)}{\theta(u)}\,du + o(1).$$

Presented to the Society, January 1, 1941; received by the editors January 23, 1941.

(1) Ahlfors [1, p. 10 and p. 16]. The number in the brackets refers to the author's paper quoted in the bibliography.

(2) For a justification of this notation see §1 (b).

If now the integral $\int_{u_0}^{\infty}[\theta'^2(u)/\theta(u)]du$ *converges*, I and II together yield an *asymptotic expression* for the difference $X(w_2)-X(w_1)$:

$$X(w_2) - X(w_1) = \pi \int_{u_1}^{u_2} \frac{1 + \psi'^2(u)}{\theta(u)} du + o(1) \qquad \text{as } u_1, u_2 \to +\infty.$$

Combining this with a result on $Y(w)$ established in this paper under the same hypothesis as in I, we obtain the following asymptotic representation for $Z(w)$:

$$Z(w) = \lambda + \pi \int_{u_0}^{u} \frac{1 + \psi'^2(t)}{\theta(t)} dt + i\pi \frac{v - \psi(u)}{\theta(u)} + o(1), \qquad \text{as } u \to +\infty,$$

uniformly with respect to v. Here λ *is a real constant.*

As to $Z'(w)$, we find under the same hypothesis as in I,

$$Z'(w) \sim \frac{\pi}{\theta(u)} \qquad \text{as } u \to +\infty,$$

uniformly in any subregion S_β: $\{(|v-\psi(u)|)/\theta(u) \leqq \beta/\pi\}$ *where* $0<\beta<\pi/2$. The "approach of u to $+\infty$ in any S_β" is the analogue of the "approach within any angle" in the case of a finite boundary point at which the boundary curve possesses a tangent. In this connection we obtain an extension of Carathéodory's well known theorem which states that the map of a region bounded by a Jordan curve onto a circle is quasi-conformal at a boundary point which is the vertex of a corner (cf. §16).

Similar expressions for $Z(w)$ and $Z'(w)$ are obtained when $\gamma \neq 0$, but $|\gamma| < \pi/2$. Further theorems are derived as corollaries from the above mentioned results.

An important part in the proof of some of these results is played by a theorem of A. Ostrowski (cf. §2) which deals with the argument of the derivative of the mapping function in a neighborhood of a point at which the boundary curve has a cusp. In the present paper a new proof of this theorem is given.

By the use of suitable transformations these results can be applied to the study of the mapping function in a neighborhood of a boundary point for various boundary configurations. Let Γ be a closed Jordan curve in the ω-plane, R the interior of Γ, let $\zeta=\zeta(\omega)$ map R conformally onto the circle $|\zeta-1|<1$, let ω_0 be a point on Γ and let $\zeta(\omega_0)=0$. By means of simple logarithmic transformations, asymptotic expressions for $\zeta(\omega)$ and $\zeta'(\omega)$ as $\omega\to\omega_0$ are derived from the above stated results. In particular, when Γ possesses a tangent at ω_0, the expression for $\zeta(\omega)$ yields a new criterion for the existence of the derivative of $\zeta(\omega)$ at ω_0 (i.e., $\lim_{\omega\to\omega_0}\zeta(\omega)/(\omega-\omega_0)$ for unrestricted approach). Another case is that in which Γ has a cusp at ω_0. It was partly for the purpose of treating this case that the present investigation was started.

In addition to the asymptotic expressions for $\zeta(\omega)$ and $\zeta'(\omega)$ as $\omega \to \omega_0$ we obtain some extensions of earlier results on the cusp which are due to Ostrowski.

Other applications include the study of $\zeta(\omega)$ for a region R whose boundary contains two "concurrent" spirals. These spirals may both approach a single "asymptotic" point ω_0, or else the set of their limiting points might be a more general point set which forms a prime end of R.

I. PRELIMINARIES

1. **Definitions.** We begin with a few definitions which will be used in the paper.

(a) SIMPLE JORDAN STRIP. *Let C_+ and C_- be two curves in the w-plane $(w=u+iv)$ represented by continuous functions*

$$v = \phi_+(u), \quad v = \phi_-(u), \qquad u \geqq u_0, \quad \phi_+(u) > \phi_-(u),$$

respectively, and let C_1 be a Jordan arc[2] *which lies in the half-plane $u \leqq u_0$ and connects the finite end points of C_+ and C_-. The curve C, consisting of C_+, C_- and C_1 decomposes the complex plane into two regions (by Jordan's theorem). Let S be the one which contains the region*

$$\phi_-(u) < v < \phi_+(u), \qquad u_0 < u < +\infty.$$

We call S a simple Jordan strip.

We set $\theta(u) = \phi_+(u) - \phi_-(u)$ and denote by θ_u the segment $\{\Re w = u, \phi_-(u) \leqq v \leqq \phi_+(u)\}$.

(b) L-TANGENT AT $u=+\infty$; L-STRIP. *Let C be a curve represented by the equation $v=\phi(u)$ where $\phi(u)$ is continuous for $u_0 \leqq u < +\infty$. We say that C has an L-tangent with the angle of inclination γ, $-\pi/2 \leqq \gamma \leqq \pi/2$ at $u=+\infty$, if the angle of inclination of any chord w_1w_2 of C $(w_1=u_1+iv_1,\ w_2=u_2+iv_2)$, $u_1<u_2$, approaches the limit γ when u_1 and u_2 approach $+\infty$ simultaneously, or, in other words, if for every $\epsilon>0$ there exists an $R(\epsilon)>0$ such that for the principal value of the argument,*

$$|\arg (w_2 - w_1) - \gamma| < \epsilon \qquad \textit{when } u_2 > u_1 > R(\epsilon).$$

Let S be a simple Jordan strip whose boundary curves C_+ and C_- have both L-tangents at $u=+\infty$ with the angles of inclination γ_+ and γ_-, $|\gamma_+| < \pi/2$, $|\gamma_-| < \pi/2$, respectively. Then, for sufficiently large u, say $u \geqq u_1 \geqq u_0$, $\theta(u)$ has bounded difference quotients and hence $\theta'(u)$ exists for $u \geqq u_1$, except possibly for a set of measure zero. Furthermore, $\theta'(u)$ is bounded and $\theta(b)-\theta(c)=\int_c^b \theta'(u)du$, $b \geqq c \geqq u_1$.

If, in particular, the L-tangents of C_+ and C_- at $u=+\infty$ have the same angle of inclination γ, then we call S an L-strip with the boundary inclination γ at $u=+\infty$.

---

[2] C_1 may pass through the infinite point ("$u=-\infty$").

The definition of an L-tangent at $u=+\infty$ is patterned after that of an L-tangent to a curve at a finite point(4). Let β be a Jordan arc which possesses a tangent at one of its end points, P. If the angle of inclination of every chord P_1P_2 of β ($P_1 \neq P_2$), approaches that of the tangent at P as P_1 and P_2 approach P simultaneously, then we say that β has an L-tangent at P.

2. **A theorem of Ostrowski on L-cusps.** Let Γ be a closed Jordan curve. If, in a neighborhood of a point P of Γ, Γ consists of two branches Γ_+ and Γ_- each of which possesses a tangent at P, and if the interior angle made by these two tangents is θ, $0<\theta\leq 2\pi$, then we shall say that Γ has a *corner of measure* θ at P. The limiting case where $\theta=0$ will be called a *cusp*. If both tangents at P are L-tangents then we shall say that Γ has an *L-corner* ($\theta>0$) or an *L-cusp* ($\theta=0$) at P.

The following theorem is due to A. Ostrowski(5).

THEOREM I. (Ostrowski.) *Let Γ be a closed Jordan curve in the ω-plane which has an L-cusp at the point P of Γ, and let $\omega=f(\zeta)$ map the circle $|\zeta|<1$ conformally onto the interior of Γ in such a manner that $\zeta=1$ corresponds to P. Then, for any determination of the argument in $|\zeta|<1$, $\lim_{\zeta\to 1}\arg\,[f'(\zeta)(\zeta-1)]$ exists for unrestricted approach.*

This theorem is an extension of a theorem by Lindelöf(6) on $\arg f'(\zeta)$ in the case that Γ has an L-corner of measure $\theta=\pi$ at P, and was first stated and proved by Ostrowski(5). We give here a proof of this theorem, which is different from the one by Ostrowski. It follows to some extent the ideas which Lindelöf used in proving his theorem.

Proof of Theorem I. (i) It may be assumed that P is the point $\omega=0$, that Γ_- follows Γ_+ when Γ is traversed in the mathematically positive direction and that the L-tangents to Γ_+ and Γ_- at $\omega=0$ fall on the negative real axis. Then we have, for a proper choice of the determination of the argument for ω_1 and ω_2 on Γ_-,

$$\lim_{\omega_1,\omega_2\to 0}\arg(\omega_2-\omega_1)=\pi \qquad (\omega_1 \text{ between } \omega=0 \text{ and } \omega_2) \tag{2.1}$$

and for ω_1 and ω_2 on Γ_+,

$$\lim_{\omega_1,\omega_2\to 0}\arg(\omega_2-\omega_1)=2\pi \qquad (\omega_2 \text{ between } \omega_1 \text{ and } \omega=0). \tag{2.2}$$

Let γ_+ and γ_- be subarcs of Γ_+ and Γ_-, respectively, such that both have $\omega=0$ as an end point and that for all ω_1, ω_2 on γ_-

$$|\arg(\omega_2-\omega_1)-\pi|<\pi/8 \qquad (\omega_1 \text{ between } \omega_2 \text{ and } \omega=0) \tag{2.3}$$

(4) The idea of an L-tangent at a finite point was introduced by Lindelöf, [1, pp. 89–91], the term "L-tangent" by Ostrowski [1, p. 93].

(5) Ostrowski [1, pp. 181–183].

(6) Lindelöf [1, pp. 89–91].

and for all ω_1, ω_2 on γ_+,

$$\left| \arg(\omega_2 - \omega_1) - 2\pi \right| < \pi/8 \qquad (\omega_2 \text{ between } \omega = 0 \text{ and } \omega_1). \tag{2.4}$$

It follows from (2.3) and (2.4) that γ_- and γ_+ can be represented in the form $(\omega = \xi + i\eta)$:

$$\eta = g_-(\xi), \qquad \eta = g_+(\xi),$$

respectively. Both of these functions are continuous and have derivatives almost everywhere in a certain interval $\xi_0 \leqq \xi \leqq 0$ $(\xi_0 < 0)$. We can assume both functions have a derivative at ξ_0, so that γ_- and γ_+ have tangents at the points P_1 $(\xi_0, g_-(\xi_0))$ and P_2 $(\xi_0, g_+(\xi_0))$, respectively. Denote the arcs OP_1 and OP_2 by Γ_-' and Γ_+', respectively.

Let γ be an arc with continuously turning tangent which joins P_1 and P_2, which has the same tangents at P_1 and P_2 as Γ, and which lies in the interior R of Γ except for its end points. Call Γ_1 the closed Jordan curve formed by γ, Γ_-' and Γ_+'. We shall prove the theorem first for the function $\omega = f_1(\zeta)$ which maps the circle $|\zeta| < 1$ onto the interior R_1 of Γ_1, and for which $f_1(1) = 0$.

(ii) We establish first an auxiliary inequality. *Let a be a point on Γ_+' and b a point on Γ_-' (so that $\Re a \geqq \xi_0$, $\Re b \geqq \xi_0$). If that determination of the argument is chosen which lies between $-\pi/2$ and $3\pi/2$ (inclusive),*

$$-\frac{\pi}{8} \leqq \arg(b - a) \leqq \pi + \frac{\pi}{8}\cdot \tag{2.5}$$

Assume first that $\Re b > \Re a$. Then only the left-hand side of (2.5) needs proof, since then $\arg(b-a) \leqq \pi/2$. Draw the line $\xi = \Re a$. This line intersects Γ_-' in a point a' whose ordinate is greater than that of a. By use of elementary properties of the angles of a triangle, it is easily seen that

$$\arg(b - a) \geqq \arg(b - a'), \qquad \arg(b - a') \geqq -\pi/8,$$

because of (2.3), since b and a' are both on Γ_-' and b is between a' and 0. The case $\Re b < \Re a$ is treated similarly. If $\Re a = \Re b$, $\arg(b-a) = \pi/2$ and (2.5) is true.

(iii) Let $f_1(e^{i\theta_1})$ and $f_1(e^{i\theta_2})$ be *interior* points of Γ_-' and Γ_+', respectively, $0 < \theta_1 < \theta_2 < 2\pi$, and let δ_0 be a positive number which is less than $\min\{\theta_1, 2\pi - \theta_2\}$ and also so small that (2.3) and (2.4) are satisfied on the arcs of Γ_1 corresponding to $0 \leqq \theta \leqq \theta_1 + \delta_0$ and $\theta_2 - \delta_0 \leqq \theta \leqq 2\pi$ of $|\zeta| = 1$, respectively. Since $\omega = f_1(\zeta)$ is continuous and univalent on $|\zeta| = 1$,

$$Q(\theta; \delta) \equiv \arg\left[f_1(e^{i(\theta+\delta)}) - f_1(e^{i\theta})\right], \quad 0 < \delta \leqq \delta_0, \ \delta \text{ fixed},$$

is continuous for all real θ, once the determination for one value, say $\theta = 0$, is selected. Since (2.3) holds on Γ_-', $Q(\theta; \delta)$ can be determined by the condition $|Q(0; \delta) - \pi| < \pi/8$. We have then also

$$\left| Q(\theta; \delta) - \pi \right| < \pi/8 \qquad \text{for } 0 \leqq \theta \leqq \theta_1. \tag{2.6}$$

We assert now that

$$(2.7)\qquad |Q(\theta;\delta)-2\pi|<\pi/8 \qquad \text{for } \theta_2-\delta\leqq\theta\leqq 2\pi-\delta.$$

To prove this it is sufficient to show that

$$(2.8)\qquad |Q(\theta_2;\delta)-2\pi|<\pi/8.$$

For since (2.4) holds on the arc of Γ_1 which corresponds to the arc $\theta_2-\delta_0\leqq\theta\leqq 2\pi$ of $|\zeta|=1$ and $Q(\theta;\delta)$ is continuous in θ, (2.7) will then follow.

To prove (2.8) we observe first that for any θ,

$$(2.9)\qquad Q(\theta+2\pi;\delta)-Q(\theta;\delta)=2\pi.$$

This is an application of the principle of the argument, since the function $f_1(\zeta e^{i\delta})-f_1(\zeta)$ is regular in $|\zeta|<1$, continuous in $|\zeta|\leqq 1$, and has exactly one zero, $\zeta=0$, in $|\zeta|\leqq 1$.

Now let $Q(\theta_2;\delta)=h_2+2k\pi$ where k is an integer and $|h_2|<\pi/8$ (because of (2.4)). Since $Q(\theta;\delta)$ is continuous, it follows from (2.4) and (2.5) that for $\theta_2\leqq\theta\leqq 2\pi$

$$(2.10)\qquad 2k\pi-\pi/8\leqq Q(\theta;\delta)\leqq 2k\pi+\pi+\pi/8.$$

Let $Q(0;\delta)=h_1+\pi$, $|h_1|<\pi/8$ by (2.6). Then by (2.9)

$$Q(2\pi;\delta)=h_1+3\pi$$

and hence by (2.10) for $\theta=2\pi$

$$2k\pi-\pi/8\leqq h_1+3\pi\leqq 2k\pi+\pi+\pi/8.$$

The left-hand side of this inequality implies $k\leqq 1$, the right-hand side $k\geqq 1$, so that $k=1$ and (2.8) is proved.

If we set $\omega_1=f_1(e^{i\theta})$, $\omega_2=f_1(e^{i(\theta+\delta)})$ then (2.6) shows that $Q(\theta;\delta)$ for $0\leqq\theta\leqq\theta_1$ is identical with the determination of arg $(\omega_2-\omega_1)$ chosen in (2.1). Hence, by (2.1), $Q(\theta;\delta)\to\pi$ *as* θ *and* δ *approach* 0 *simultaneously*. Similarly, we infer from (2.7) and (2.2) that $Q(\theta;\delta)\to 2\pi$ *as* $\theta\to 2\pi$ *and* $\delta\to 0$ *with the restriction that* $\theta\leqq 2\pi-\delta$.

Finally we note that there exists an $M>0$ such that for all $0\leqq\theta\leqq 2\pi$ and all $0<\delta\leqq\delta_0$

$$(2.11)\qquad |Q(\theta;\delta)|\leqq M.$$

For $0\leqq\theta\leqq\theta_1$ this follows from (2.6) for $\theta_2-\delta\leqq\theta\leqq 2\pi-\delta$ from (2.7), for $2\pi-\delta\leqq\theta\leqq 2\pi$ from (2.10) with $k=1$, and for $\theta_1\leqq\theta\leqq\theta_2-\delta$ from (2.3), (2.4) and the fact that γ has a continuously turning tangent.

(iv) For fixed $\delta\leqq\delta_0$, the function

$$g(\zeta;\delta)\equiv\frac{f_1(\zeta e^{i\delta})-f_1(\zeta)}{\zeta(e^{i\delta}-1)}\quad \text{for } \zeta\neq 0,\ g(0;\delta)=f_1'(0),$$

is regular in $|\zeta|<1$, continuous in $|\zeta|\leqq 1$, and never vanishes because $f_1(\zeta)$ is univalent in $|\zeta|\leqq 1$. Therefore, any branch of arg $g(\zeta;\delta)$ is harmonic in $|\zeta|<1$ and continuous in $|\zeta|\leqq 1$. We choose that branch of arg $g(\zeta;\delta)$ which reduces to

$$Q(\theta;\delta)-\delta/2-\theta-\pi/2, \qquad 0\leqq\theta\leqq 2\pi,$$

for $|\zeta|=1$. Consider now, for $|\zeta|<1$, the harmonic function

$$P(\zeta;\delta)=\arg\,[g(\zeta;\delta)(\zeta-1)]=\arg g(\zeta;\delta)+\arg(\zeta-1), \tag{2.12}$$

where arg $(\zeta-1)$ is determined by the condition that it reduces to π for $\zeta=0$. $P(\zeta;\delta)$ is bounded in $|\zeta|<1$ and it has continuous boundary values on $|\zeta|=1$ except at $\zeta=1$. Hence it can be represented by the Poisson integral

$$P(\zeta;\delta)=\frac{1}{2\pi}\int_0^{2\pi}P(e^{i\theta};\delta)\frac{(1-\rho^2)d\theta}{1+\rho^2-2\rho\cos(\theta-\alpha)}, \qquad \zeta=\rho e^{i\alpha},\quad \rho<1.$$

(v) Let $\epsilon>0$ be given. Then by our statement at the end of section (iii) there exists an $\eta=\eta(\epsilon)<\epsilon/4$ such that, for all positive $\delta\leqq\min(\delta_0,\eta)$:

$$|Q(\theta;\delta)-\pi|<\epsilon/4 \qquad \text{for } 0\leqq\theta\leqq\eta \tag{2.13}$$

and

$$|Q(\theta;\delta)-2\pi|<\epsilon/4 \qquad \text{for } 2\pi-\eta\leqq\theta\leqq 2\pi-\delta. \tag{2.14}$$

Observing that, for the determination of arg $(\zeta-1)$ selected in (2.12)

$$\arg(e^{i\theta}-1)=\frac{\theta}{2}+\frac{\pi}{2}, \qquad 0<\theta<2\pi,$$

we find, for $0<\theta<\eta$,

$$\begin{aligned}|P(e^{i\theta};\delta)-\pi| &= \left|Q(\theta;\delta)-\frac{\delta}{2}-\theta-\frac{\pi}{2}+\left(\frac{\theta}{2}+\frac{\pi}{2}\right)-\pi\right|\\ &\leqq |Q(\theta;\delta)-\pi|+\frac{\delta}{2}+\frac{|\theta|}{2}\\ &\leqq |Q(\theta;\delta)-\pi|+\frac{\delta}{2}+\frac{\eta}{2},\end{aligned}$$

and for $2\pi-\eta\leqq\theta\leqq 2\pi-\delta$,

$$\begin{aligned}|P(e^{i\theta};\delta)-\pi| &\leqq |Q(\theta;\delta)-2\pi|+\frac{\delta}{2}+\left|\frac{\theta}{2}-\pi\right|\\ &\leqq |Q(\theta;\delta)-2\pi|+\frac{\delta}{2}+\frac{\eta}{2}\cdot\end{aligned}$$

Because of (2.13) and (2.14), we have therefore

$$(2.15) \qquad |P(e^{i\theta};\delta)-\pi| < \frac{\epsilon}{4}+\frac{\epsilon}{8}+\frac{\epsilon}{8}=\frac{\epsilon}{2}$$

$$\text{for } 0 \leqq \theta \leqq \eta \text{ and } 2\pi-\eta \leqq \theta \leqq 2\pi-\delta.$$

To estimate the difference $|P(\zeta;\delta)-\pi|$ we decompose the Poisson integral as follows:

$$P(\zeta;\delta)=\frac{1}{2\pi}\left\{\int_0^{\eta}+\int_{\eta}^{2\pi-\eta}+\int_{2\pi-\eta}^{2\pi-\delta}+\int_{2\pi-\delta}^{2\pi}\right\}.$$

Then we find, from (2.15), (2.12), and (2.11),

$$\begin{aligned}|P(\zeta;\delta)-\pi| \leqq\ & \frac{1}{2\pi}\frac{\epsilon}{2}\int_0^{\eta}\frac{(1-\rho^2)d\theta}{1+\rho^2-2\rho\cos(\theta-\alpha)} \\ & +\frac{M+2\pi}{2\pi}\frac{2\pi(1-\rho^2)}{1+\rho^2-2\rho\cos(\eta-|\alpha|)} \\ & +\frac{1}{2\pi}\frac{\epsilon}{2}\int_{2\pi-\eta}^{2\pi-\delta}\frac{(1-\rho^2)d\theta}{1+\rho^2-2\rho\cos(\theta-\alpha)} \\ & +\frac{M+2\pi}{2\pi}\cdot\frac{\delta(1-\rho^2)}{(1-\rho)^2}.\end{aligned}$$

Now we keep ζ fixed and let $\delta\to 0$. Since M is independent of δ, we find

$$|\arg[f_1'(\zeta)(\zeta-1)]-\pi| \leqq \epsilon+(M+2\pi)\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\eta-|\alpha|)}.$$

Hence, as $\zeta\to 1$,

$$(2.16) \qquad \limsup|\arg[f_1'(\zeta)(\zeta-1)]-\pi| \leqq \epsilon,$$

q.e.d.

(vi) It remains to prove the theorem for the original function $f(\zeta)$. The inverse function of $\omega=f(\zeta)$ maps R_1 onto a region E of the circle $|\zeta|<1$ and the arc P_1OP_2 of Γ_1 onto an arc of the circumference $|\zeta|=1$ which contains $\zeta=1$ as an interior point. If $\zeta=\zeta(s)$, $\zeta(1)=1$, is a suitably chosen function which maps the circle $|s|<1$ onto E, then $f_1(s)=f(\zeta(s))$. By Schwarz's reflection principle, $\zeta(s)$ is regular at $\zeta=1$ and $\zeta'(1)\neq 0$. Now we have, in a sufficiently small neighborhood of $\zeta=1$, $|\zeta|<1$,

$$\arg f'(\zeta)=\arg f_1'(s)-\arg\zeta'(s)$$

and

$$\arg[f'(\zeta)(\zeta-1)]=\arg[f_1'(s)(s-1)]-\arg\zeta'(s)+\arg\frac{\zeta-1}{s-1}.$$

Letting $\zeta \to 1$ in $|\zeta| < 1$ and consequently $s \to 1$ in $|s| < 1$, we find the desired result from (2.16).

3. **An application.** As an application of Theorem I we prove the following

THEOREM II. *Let S be an L-strip with the boundary inclination $\gamma = 0$ at $u = +\infty$. Let $w = W(z)$, $z = x + iy$, map the strip $|y| < \pi/2$ conformally onto S in such a manner that $\Re W(x+iy) \to +\infty$ as $x \to +\infty$. Then*

(a) *For a suitable choice of the branch of the argument,* $\lim_{z\to+\infty} \arg W'(z) = 0$, *for unrestricted approach.*

(b) *For z_1 and z_2 in any fixed strip $|y| \leq \beta < \pi/2$ which satisfy the condition $|z_2 - z_1| \leq M$ (M a constant), $\lim [W'(z_2)/W'(z_1)] = 1$ as $x_1 = \Re z_1$ (and hence $x_2 = \Re z_2$) approaches $+\infty$, uniformly in $|y| \leq \beta$*(7).

Proof. (a) The transformation $\omega = 1/w$ maps S onto a region R bounded by a closed Jordan curve(8) Γ in such a manner that $u = +\infty$ corresponds to $\omega = 0$. Let w_1 and w_2 be two points on one of the boundary curves C_+ or C_- of S and $\omega_1 = 1/w_1$ and $\omega_2 = 1/w_2$ their images on Γ. Since S is an L-strip with the boundary inclination $\gamma = 0$ at $u = +\infty$, the principal value of $\arg w$ $(-\pi < \arg w \leq \pi)$ is single-valued in S *if $u = \Re w$ is sufficiently large*, and $\arg w_1$ and $\arg w_2$ approach 0 as w_1 and w_2 approach ∞ along C_+ or C_-. Hence, it follows from the relation $\arg(\omega_2 - \omega_1) = \arg(w_1 - w_2) - \arg w_1 - \arg w_2$, which holds when the principal value is taken for each of the arguments, provided $\Re w_1 > \Re w_2$ and both are sufficiently large, that Γ has an L-cusp at $\omega = 0$.

Let now, for $|\zeta| < 1$, $z = \log[(1+\zeta)/(1-\zeta)]$ be the branch of the logarithm which is 0 when $\zeta = 0$. The function

$$\omega = f(\zeta) = \frac{1}{W(z)}, \qquad z = \log \frac{1+\zeta}{1-\zeta}$$

maps the circle $|\zeta| < 1$ conformally onto R. It can be defined as a continuous function in $|\zeta| \leq 1$, and then $f(1) = 0$. Hence by Theorem I,

$$\lim_{\zeta \to 1} \arg [f'(\zeta)(\zeta - 1)] = l$$

exists, or

$$\lim_{z \to +\infty} \arg W'(z) = \lim_{\zeta \to 1} \arg \left[(W(z))^2 f'(\zeta) \frac{\zeta^2 - 1}{2} \right] = l$$

exists. That this limit is 0 for a suitable choice of the branch of the argument can be seen in the following way. The image L in the w-plane of the real axis

(7) Part (b) is due to Ostrowski [1, p. 185, relation (68.6), and p. 177, relation (64.2)].

(8) If the arc C_1 of the boundary curve of S passes through $w = \infty$, Γ will have a double point at $\omega = 0$. In this case, however, Γ will be a Jordan curve on the Riemann surface of $(\omega - a)^{1/2}$ for a suitable choice of the point a $(a \neq 0)$.

in the z-plane by means of $w=W(z)$ has a tangent at the point $W(x)$ which forms the angle arg $W'(x)$ with the positive u-axis. If $l\neq 0$, we may assume that l is also not a multiple of 2π. Since L lies in S, the slope of any chord through two points $W(x_0)$ and $W(x_1)$, x_0 *fixed*, will approach 0 as $x_1\to+\infty$, no matter how large x_0 might be chosen. On the other hand, $\lim_{x\to+\infty}\arg W'(z)=l$ implies that L has an L-tangent with the direction l at $u=+\infty$, which is impossible unless $l=0$ or a multiple of 2π.

(b) To prove part (b) we note first: If $F(z)=U(z)+iV(z)$ is regular for $|y|<\pi/2$ and if $\lim_{x\to+\infty} V(z)=V_0$ in $|y|<\pi/2$, then uniformly in any fixed sub-strip $|y|\leqq\beta<\pi/2$, $\lim_{x\to+\infty} F'(z)=0$(9).

This statement follows immediately from the integral representation(10):

$$F'(z)=\frac{i}{\pi r}\int_0^{2\pi} V(z+re^{i\theta})e^{-i\theta}d\theta=\frac{i}{\pi r}\int_0^{2\pi}[V(z+re^{i\theta})-V_0]e^{-i\theta}d\theta,$$

where $0<r<\pi/2-\beta$, $z=x+iy$, $|y|\leqq\beta$.

Now, if z_1 and z_2 are points in $|y|\leqq\beta<\pi/2$, and $|z_2-z_1|\leqq M$,

$$|F(z_2)-F(z_1)|=\left|\int_{z_1}^{z_2}F'(\zeta)d\zeta\right|=o(|z_2-z_1|)=o(1)$$

as $x_1\to\infty$, uniformly in $|y|\leqq\beta$. Applying this result to

$$F(z)\equiv\log W'(z)=\log|W'(z)|+i\arg W'(z)$$

and using the result of part (a), we have, for $|z_2-z_1|\leqq M$,

$$\lim_{z_1\to+\infty}[\log W'(z_2)-\log W'(z_1)]=0$$

or

$$\lim_{z_1\to+\infty}\frac{W'(z_2)}{W'(z_1)}=1,$$

q.e.d.

4. **Some auxiliary results.** As an application of Theorem II we prove the following

LEMMA 1. *Let S be an L-strip in the w-plane with the boundary inclination* 0 *at $u=+\infty$. For all u exceeding a certain number u_1, let l_u denote a line segment within S which joins the point $w=u+i\phi_-(u)$ of C_- with some point of C_+ and which forms the angle $\gamma(u)$, $\lim_{u\to+\infty}\gamma(u)=0$, with the positive v-axis*(11). *Let*

(9) This result is well known; see for example, Wolff [1, p. 221, §6], Ostrowski [2, p. 23, Theorem V, Part 3]. For the last part of this proof compare Ostrowski [2, p. 31, Theorem VII].

(10) Copson [1, p. 88, Example 1].

(11) The angle γ which a line forms with the positive v-axis is the smaller of the two angles between them (if there be a smaller one), and it is considered as positive if the direction of rotation from the positive v-axis to the line is counterclockwise.

$z=Z(w)=X(w)+iY(w)$ map the strip S conformally onto the strip $|y|<\pi/2$ in such a manner that $\lim_{u\to+\infty} X(w)=+\infty$. If

$$x^* = \max_{w\in l_u} X(w), \qquad x_* = \min_{w\in l_u} X(w), \qquad m(u) = \max_{w\in l_u} |\arg Z'(w)|,$$

then, for all sufficiently large u,

$$x^* - x_* \leqq 2\pi \tan \{m(u) + |\gamma(u)|\} \to 0 \qquad \text{as } u \to +\infty.$$

Proof. By Theorem I, a suitable branch of $\arg Z'(w)$ approaches 0 uniformly in S as $u\to+\infty$. Let, for $u\geqq u_2\geqq u_1$, $|\arg Z'(w)|<\pi/8$, $|\gamma(u)|<\pi/8$. The image of l_u in the z-plane by means of $z=Z(w)$, is an arc λ_u joining $z^*=Z(u+i\phi_-(u))$ on $y=-\pi/2$ with a point on $y=\pi/2$. If $w\in l_u$ (w in the interior of S), then λ_u has a tangent at $z=Z(w)$ which forms the angle $\tau=\arg Z'(w)+\gamma(u)$ with the positive y-axis. A simple application of the mean value theorem shows that no point of λ_u is in the exterior of the isosceles triangle whose top is at z^*, whose base is on $y=\pi/2$ and whose angle at z^* is $2[m(u)+|\gamma(u)|]$. The base of this triangle is $2\pi \tan [m(u)+|\gamma(u)|]$, and therefore $x^*-x_*\leqq 2\pi \tan [m(u)+|\gamma(u)|]$.

THEOREM III. *Let S be a simple Jordan strip and let $z=Z(w)=X(w)+iY(w)$ map S conformally onto the strip $|y|<\pi/2$ in such a manner that $\lim_{u\to+\infty} X(w)=+\infty$. Let $w_1=u_1+iv_1$, $w_2=u_2+iv_2$, $u_0\leqq u_1\leqq u_2$, $x_1=X(w_1)$, $x_2=X(w_2)$. Then*

(a) *the integral*

$$\pi\int_{u_1}^{u_2} \frac{du}{\theta(u)} \leqq x_2 - x_1 + 4\pi; \tag{4.1}$$

(b) *if S is an L-strip with the boundary inclination 0 at $u=+\infty$,*

$$\pi\int_{u_1}^{u_2} \frac{du}{\theta(u)} \leqq x_2 - x_1 + o(1) \qquad \text{as } u \to +\infty, \tag{4.2}$$

uniformly with respect to v_1, v_2.

Proof. (a) Part (a) is a theorem of Ahlfors (see Ahlfors [1, p. 10] or Nevanlinna [1, p. 92]).

(b) In the proof of (4.1) an inequality (Ahlfors [1, p. 8, relation (3) or Nevanlinna [1, p. 90, relation (35)]) is first derived which contains the following

$$\pi\int_{u_1}^{u_2} \frac{du}{\theta(u)} \leqq x_2^* - x_{*1},$$

where $x_{*1}=\min_{w\in\theta_u} X(w)$, $x_2^*=\max_{w\in\theta_u} X(w)$. Now, by Lemma 1, (applied with $l_u\equiv\theta_u$), $x_2^*-x_2\to 0$ and $x_1-x_{*1}\to 0$ as $u_1\to+\infty$. Hence, if we write $x_2^*=x_2+(x_2^*-x_2)$, $x_{*1}=x_1+(x_{*1}-x_1)$, the result (4.2) follows immediately.

II. The first basic inequality

5. Statement of Theorem IV. The content of Theorem IV (b) will be referred to as the first basic inequality.

THEOREM IV. *Let S be a simple Jordan strip and let the functions $v=\phi_+(u)$ and $v=\phi_-(u)$ representing the boundary curves C_+ and C_- of S have uniformly bounded difference quotients for $u\geqq u_0$. Let*

$$\theta(u) = \phi_+(u) - \phi_-(u), \qquad \psi(u) = \tfrac{1}{2}[\phi_+(u) + \phi_-(u)].$$

Suppose that the function $w=W(z)=U(z)+iV(z)$, $z=x+iy$, maps the strip $|y|<\pi/2$ conformally onto S in such a manner that $x=+\infty$ corresponds to $u=+\infty$. Let $z_1=x_1+iy_1$ and $z_2=x_2+iy_2$ be two points in $|y|<\pi/2$, $x_1<x_2$, and let $u_1=U(z_1)$, $u_2=U(z_2)$.

(a) *If $|\phi_+'(u)|$ and $|\phi_-'(u)|\leqq m$ for $u\geqq u_0$, there is an x_0, depending on u_0, such that for $x_0\leqq x_1\leqq x_2$*

$$x_2 - x_1 \leqq \pi\int_{u_1}^{u_2}\frac{1+\psi'^2(u)}{\theta(u)}du + \frac{\pi}{12}\int_{u_1}^{u_2}\frac{\theta'^2(u)}{\theta(u)}du + 8\pi(1+\tfrac{4}{3}m^2). \tag{5.1}$$

(b) *If S is an L-strip with the boundary inclination $\gamma=0$ at $u=+\infty$, then*

$$x_2 - x_1 \leqq \pi\int_{u_1}^{u_2}\frac{1+\psi'^2(u)}{\theta(u)}du + \frac{\pi}{12}\int_{u_1}^{u_2}\frac{\theta'^2(u)}{\theta(u)}du + o(1) \tag{5.2}$$

as $x_1\to+\infty$, uniformly with respect to y_1, y_2.

REMARK. Ahlfors proves [1, pp. 12–16] the following theorem. Let S be a region represented in the form $-\theta(u)/2<v<\theta(u)/2$, $-\infty<u<+\infty$, where $\theta(u)$ is of bounded variation in any finite interval and $0<\theta(u)<L$ for all u. Then, if x_1, x_2, u_1, u_2 have the same meaning as in Theorem IV,

$$x_2 - x_1 \leqq \pi\int_{u_1}^{u_2}\frac{du}{\theta(u)} + 8\pi + 16\pi L^2\frac{\theta_m^2+V}{\theta_m^4}, \tag{5.3}$$

where θ_m is the infimum of $\theta(u)$ for $u_1-4L\leqq u\leqq u_2+4L$ and V is the total variation of $\theta^2(u)$ in this interval. (It should be noted that $\psi(u)\equiv 0$ here because S is symmetrical with respect to the real axis.) While the hypotheses of Theorem IV regarding the smoothness of the boundary of S are more stringent than those of Ahlfors' theorem, Theorem IV contains no restriction as to the symmetry of S or boundedness of $\theta(u)$. Moreover, since $V=2\int_{u=u_1-4L}^{u_2+4L}\theta(u)|d\theta(u)|$, for continuous $\theta(u)$,

$$16\pi L^2\frac{\theta_m^2+V}{\theta_m^4} \geqq 16\pi\frac{L^2}{\theta_m^2}\left\{1+2\int_{u=u_1-4L}^{u=u_2+4L}\frac{|d\theta(u)|}{\theta(u)}\right\}, \tag{5.4}$$

and this shows that under the hypotheses of Theorem IV (b) the "remainder" term $\pi\int_{u_1}^{u_2}(\theta'^2(u)/\theta(u))du+o(1)$ is smaller than the one in (5.3) if u_1 is sufficiently large. This becomes important in the case when $\liminf_{u\to+\infty}\theta(u)=0$, since then (5.4) is not bounded as $u_2\to+\infty$ (u_1 fixed), while $\int_{u_1}^{\infty}(\theta'^2(u)/\theta(u))du$ converges for very general classes of functions $\theta(u)$.—In the proof of Theorem IV (as well as in that of Theorem VI below) we make use of Ahlfors' method of relating area and arc length employed in the proof of his inequalities.

6. **Proof of Theorem** IV. 1. Let for $x_1\leqq x\leqq x_2$, $a<U(z)<b$. We assume first that the functions $\phi_+(u)$ and $\phi_-(u)$ have two continuous derivatives for $a\leqq u\leqq b$.

The functions

$$(6.1)\qquad \bar{u}=h(u)=\int_{u_0}^{u}\frac{\eta dt}{(\theta(t)(1+t^2))^{1/2}},\quad \bar{v}=\frac{v-\psi(u)}{\theta(u)},\quad \eta>0,\text{ a constant,}$$

map the domain $S(a, b)$: $\{\phi_-(u)\leqq v\leqq\phi_+(u),\ a\leqq u\leqq b\}$ in a one-to-one and continuous manner onto a rectangle: $\{\bar{a}\leqq\bar{u}\leqq\bar{b},\ |\bar{v}|\leqq 1\}$ of the $\bar{w}$-plane, $\bar{w}=\bar{u}+i\bar{v}$. The function $W(z)$ maps the line segment $\{x, |y|\leqq\pi/2\}$, $x_1\leqq x\leqq x_2$, onto a rectifiable arc l_x in S and the transformation (6.1) carries l_x into a rectifiable arc $\bar{l}_x$ within the strip $|\bar{v}|\leqq 1$ of the w-plane. If the variable arc lengths of l_x and $\bar{l}_x$ are denoted by s and $\bar{s}$, respectively, the

$$\text{length of }\bar{l}_x=\int_{l_x}\frac{d\bar{s}}{ds}ds=\int_{-\pi/2}^{\pi/2}\frac{d\bar{s}}{ds}|W'(x+iy)|dy.$$

Since $\bar{l}_x$ connects a point on the line $\bar{v}=+1$ with another on $\bar{v}=-1$, its length is greater than or equal to 1. Hence,

$$(6.2)\quad 1\leqq\left\{\int_{-\pi/2}^{\pi/2}\frac{d\bar{s}}{ds}|W'(x+iy)|dy\right\}^2\leqq\pi\int_{-\pi/2}^{\pi/2}\left(\frac{d\bar{s}}{ds}\right)^2|W'(x+iy)|^2dy.$$

Now, if $u(s)$, $v(s)$ denote the parametric representation of l_x in terms of s,

$$\begin{aligned}\left(\frac{d\bar{s}}{ds}\right)^2&=\left(\frac{d\bar{u}}{du}\frac{du}{ds}\right)^2+\left(\frac{\partial\bar{v}}{\partial u}\frac{du}{ds}+\frac{\partial\bar{v}}{\partial v}\frac{dv}{ds}\right)^2\\
&=\left[h'^2(u)+\left(\frac{\partial\bar{v}}{\partial u}\right)^2\right]\left(\frac{du}{ds}\right)^2+2\frac{\partial\bar{v}}{\partial u}\frac{\partial\bar{v}}{\partial v}\frac{du}{ds}\frac{dv}{ds}+\left(\frac{\partial\bar{v}}{\partial v}\right)^2\left(\frac{dv}{ds}\right)^2\\
&\leqq\left[h'^2(u)+\left(\frac{\partial\bar{v}}{\partial u}\right)^2\right]\left(\frac{du}{ds}\right)^2\\
&\qquad+2\left|\frac{\partial\bar{v}}{\partial v}\right|\left|\frac{du}{ds}\right|\left|\frac{dv}{ds}\right|\left(h'^2(u)+\left(\frac{\partial\bar{v}}{\partial u}\right)^2\right)^{1/2}+\left(\frac{\partial\bar{v}}{\partial v}\right)^2\left(\frac{dv}{ds}\right)^2\\
&=\left\{\left(h'^2(u)+\left(\frac{\partial\bar{v}}{\partial u}\right)^2\right)^{1/2}\frac{du}{ds}+\left|\frac{\partial\bar{v}}{\partial v}\right|\left|\frac{dv}{ds}\right|\right\}^2.\end{aligned}$$

Hence, by Schwarz's inequality,

$$\left(\frac{d\bar{s}}{ds}\right)^2 \leqq \left[h'^2(u) + \left(\frac{\partial\bar{v}}{\partial u}\right)^2 + \left(\frac{\partial\bar{v}}{\partial v}\right)^2\right]\left[\left(\frac{du}{ds}\right)^2 + \left(\frac{dv}{ds}\right)^2\right]$$
$$= h'^2(u) + \left(\frac{\partial\bar{v}}{\partial u}\right)^2 + \left(\frac{\partial\bar{v}}{\partial v}\right)^2,$$

since $(du/ds)^2+(dv/ds)^2=1$. Observing that (we leave off the argument u)

$$\frac{\partial\bar{v}}{\partial u} = -\frac{\theta\psi' + (v-\psi)\theta'}{\theta^2}, \qquad \frac{\partial\bar{v}}{\partial v} = \frac{1}{\theta}$$

(ψ', θ' are derivatives with respect to u), we find at every point (u, v) of l_x

$$(6.3)\quad \left(\frac{d\bar{s}}{ds}\right)^2 \leqq \frac{\eta^2}{\theta\cdot(u^2+1)} + \frac{\theta^2\psi'^2 + 2(v-\psi)\theta\theta'\psi' + (v-\psi)^2\theta'^2}{\theta^4} + \frac{1}{\theta^2} \equiv f(u, v).$$

Hence, from (6.2)

$$1 \leqq \pi\int_{-\pi/2}^{\pi/2} f(U(z), V(z))\,|\,W'(x+iy)\,|^2 dy.$$

Integrating this inequality between the limits x_1, x_2 with respect to x, we find

$$(x_2 - x_1) \leqq \pi\int_{x_1}^{x_2} dx\int_{-\pi/2}^{\pi/2} f(U(z), V(z))\,|\,W'(x+iy)\,|^2 dy.$$

Now, introducing the variables (u, v) in place of (x, y) by means of the transformation $w=W(x+iy)$, we obtain

$$x_2 - x_1 \leqq \pi\iint_{(T)} f(u, v)\,du\,dv$$

where T is the image of the rectangle $x_1<x<x_2$, $|y|<\pi/2$, in the w-plane. If $u_{*1}=\min_{|y|\leqq\pi/2} U(x_1+iy)$ and $u_2^*=\max_{|y|\leqq\pi/2} U(x_2+iy)$, it is clear that T is contained in the region $\phi_-(u)<v<\phi_+(u)$, $u_{*1}<u<u_2^*$. Using this fact and substituting the value in (6.3) for $f(u, v)$, we find

$$x_2 - x_1 \leqq \pi\int_{u_{*1}}^{u_2^*}\int_{\phi_-(u)}^{\phi_+(u)} \frac{1+\psi'^2(u)}{\theta^2(u)}\,dv\,du + \pi\eta^2\int_{u_{*1}}^{u_2^*}\int_{\phi_-(u)}^{\phi_+(u)} \frac{dv\,du}{\theta(u)(u^2+1)}$$
$$+ \pi\int_{u_{*1}}^{u_2^*}\int_{\phi_-(u)}^{\phi_+(u)} \frac{2(v-\psi)\theta\theta'\psi' + (v-\psi)^2\theta'^2}{\theta^4}\,dv\,du.$$

The integration with respect to v can be easily carried through. The first two integrals yield the result

$$\pi \int_{u_{*1}}^{u_2^*} \frac{1 + \psi'^2(u)}{\theta(u)} du + \pi\eta^2 \int_{u_{*1}}^{u_2^*} \frac{du}{1 + u^2}.$$

The third integral is equal to

$$2\pi \int_{u_{*1}}^{u_2^*} 0 \cdot \frac{\theta'\psi'}{\theta^3} du + \pi \int_{u_{*1}}^{u_2^*} \frac{2}{3} \left(\frac{\theta}{2}\right)^3 \frac{\theta'^2}{\theta^4} du = \frac{\pi}{12} \int_{u_{*1}}^{u_2^*} \frac{\theta'^2(u)}{\theta(u)} du.$$

Hence we finally obtain

$$x_2 - x_1 \leqq \pi \int_{u_{*1}}^{u_2^*} \frac{1 + \psi'^2(u)}{\theta(u)} du + \frac{\pi}{12} \int_{u_{*1}}^{u_2^*} \frac{\theta'^2(u)}{\theta(u)} du + \pi\eta^2 \int_{-\infty}^{\infty} \frac{du}{1 + u^2}.$$

This inequality is true for every η. Keeping x_1, x_2 fixed and letting $\eta \to 0$ we find

$$(6.4) \qquad x_2 - x_1 \leqq \pi \int_{u_{*1}}^{u_2^*} \frac{1 + \psi'^2(u)}{\theta(u)} du + \frac{\pi}{12} \int_{u_{*1}}^{u_2^*} \frac{\theta'^2(u)}{\theta(u)} du.$$

2. If the hypothesis made at the beginning of part 1 of this proof (that $\phi_+(u)$ and $\phi_-(u)$ have continuous second derivatives for $a \leqq u \leqq b$) is not satisfied, we replace the arcs $\beta_+ : v = \phi_+(u)$ and $\beta_- : v = \phi_-(u)$, $a \leqq u \leqq b'$ $(b' > b)$ by certain arcs $\beta_+^{(n)}$ and $\beta_-^{(n)}$ for which this assumption is true and which converge to β_+ and β_-, respectively, as $n \to +\infty$. We proceed as follows:

Since $\phi_+(u)$ and $\phi_-(u)$ have bounded difference quotients, $\phi_+'(u)$ and $\phi_-'(u)$ exist almost everywhere in $a \leqq u \leqq b'$ and are, in absolute value, less than or equal to m. There exist therefore two sequences of continuous functions, $\phi_+'(u; n)$ and $\phi_-'(u; n)$, $a \leqq u \leqq b$, such that

$$(6.5.1) \quad \int_a^b | \phi_+'(u; n) - \phi_+'(u) | du \to 0, \int_a^b | \phi_-'(u; n) - \phi_-'(u) | du \to 0 \text{ as } n \to \infty,$$

$$(6.5.2) \quad | \phi_+'(u; n) | \leqq 2m, \quad | \phi_-'(u; n) | \leqq 2m, \qquad a \leqq u \leqq b, \; n = 1, 2, 3, \cdots,$$

$$(6.5.3) \quad \frac{d}{du} \phi_+'(u; n), \quad \frac{d}{du} \phi_-'(u; n) \qquad \text{exist and are continuous for } a \leqq u \leqq b.$$

Now we define

$$(6.6) \quad \phi_+(u; n) = \phi_+(a) + \int_a^u \phi_+'(t; n) dt, \quad \phi_-(u; n) = \phi_-(a) + \int_a^u \phi_-'(t; n) dt,$$
$$a \leqq u \leqq b.$$

Furthermore, we may assume that $\phi_+(b) \neq 0$ and $\phi_-(b) \neq 0$ and set for $b \leqq u \leqq b'$

$$q_+(u; n) = q_n + \frac{1 - q_n}{b' - b} (u - b) \quad \text{where } q_n = \frac{\phi_+(b; n)}{\phi_+(b)},$$

$$q_-(u; n) = p_n + \frac{1 - p_n}{b' - b}(u - b) \quad \text{where } p_n = \frac{\phi_-(b; n)}{\phi_-(b)},$$

and define

$$(6.7) \quad \phi_+(u; n) = \phi_+(u)q_+(u; n), \quad \phi_-(u; n) = \phi_-(u)q_-(u; n) \quad \text{for } b \leqq u \leqq b'.$$

Then it is evident from (6.5.1), (6.6) and (6.7) that *the arcs*

$$\beta_+^{(n)}: \; v = \phi_+(u; n) \quad \textit{and} \quad \beta_-^{(n)}: \; v = \phi_-(u; n), \qquad a \leqq u \leqq b',$$

converge to the arcs β_+ *and* β_-, *respectively, in the sense that*

$$(6.8) \quad \lim_{n\to\infty} \phi_+(u; n) = \phi_+(u), \quad \lim_{n\to\infty} \phi_+(u; n) = \phi_-(u), \; \textit{uniformly for } a \leqq u \leqq b'.$$

If now, for sufficiently large n, the arcs β_+ and β_- are replaced by $\beta_+^{(n)}$ and $\beta_-^{(n)}$, respectively, we obtain a Jordan strip S_n whose boundary curves satisfy the hypotheses stated in the theorem and in the beginning of part 1 of this proof. Let $W_n(z)$ be the function which maps the strip $|y| < \pi/2$ onto S_n in such a manner that $x = +\infty$ corresponds to $u = +\infty$ and that $W_n(0) = W(0)$. Then we may define $u_{*1}^{(n)}$ and $u_2^{*(n)}$ for $W_n(z)$ in the same way as u_{*1} and u_2^* are defined for $W(z)$. Since $\beta_+^{(n)}$ and $\beta_-^{(n)}$ converge to β_+ and β_-, respectively, in the above specified manner,

$$(6.9) \qquad \lim_{n\to\infty} u_{*1}^{(n)} = u_{*1}, \qquad \lim_{n\to\infty} u_2^{*(n)} = u_2^*\text{(12)}.$$

Let $\theta_n(u)$ and $\psi_n(u)$ have the same meaning for S_n as $\theta(u)$ and $\psi(u)$ for S. We now apply (6.4) to the region S_n, where n is sufficiently large, and obtain an analogous inequality in which θ, ψ', θ', u_{*1}, and u_2^* are replaced by θ_n, ψ_n', θ_n', $u_{*1}^{(n)}$, $u_2^{*(n)}$, respectively. Here we let $n \to \infty$, and it follows from (6.5.1) and (6.5.2), that the inequality holds for the original region S without the assumption of the existence of the second derivatives of $\phi_+(u)$ and $\phi_-(u)$.

---

(12) To see this, we may assume that $w = 0$ is neither in S nor on the boundary of S and transform S and S_n by means of the function $\omega = 1/w$ into limited regions R and R_n, respectively. R and R_n are bounded by closed curves Γ and Γ_n, respectively, which coincide except for the arcs $\gamma_+^{(n)}$ and $\gamma_-^{(n)}$ of Γ_n, the images of $\beta_+^{(n)}$ and $\beta_-^{(n)}$, and γ_+ and γ_- of Γ, the images of β_+ and β_-. From (6.8) it follows that the curves Γ_n converge to Γ in the sense that their Fréchet distance approaches 0. (The Fréchet distance d of two Jordan curves C and C' is defined as follows: For any continuous one-to-one transformation of C onto C', the distance of corresponding points has a maximum. The greatest lower bound of these maxima for all possible transformations is d.) If $\omega_n(\zeta)$ and $\omega(\zeta)$ map $|\zeta| < 1$ conformally onto R_n and R, respectively, and if $\omega_n(0) = \omega(0)$ and $\omega_n(1) = \omega(1) = 0$ (the image of $u = +\infty$), then it follows from a theorem of Radó [1, pp. 180–186] that $\omega_n(\zeta) \to \omega(\zeta)$ uniformly in $|\zeta| \leqq 1$ as $n \to \infty$. Now the functions $W_n(z)$ and $W(z)$ of the text are

$$W_n(z) = [\omega_n(\zeta)]^{-1}, \qquad W(z) = [\omega(\zeta)]^{-1}, \qquad \text{where } z = \log\frac{1+\zeta}{1-\zeta},$$

and it follows therefore that $W_n(z) \to W(z)$ *uniformly* in any fixed rectangle $\xi_1 \leqq x \leqq \xi_2$, $|y| \leqq \pi/2$. This implies (6.9).

3. Now the proof is easily completed. We treat the cases (a) and (b) of our theorem separately.

(a) By Theorem III (a),

$$\pi\int_{u_{*1}}^{u_1}\frac{du}{\theta(u)} \leqq x_1 - x_1 + 4\pi = 4\pi, \qquad \pi\int_{u_2}^{u_2^*}\frac{du}{\theta(u)} \leqq 4\pi.$$

Moreover, since

$$1 + \psi'^2(u) \leqq 1 + m^2, \qquad \theta'^2(u) \leqq 4m^2,$$

we find (5.1) from (6.4).

(b) In the case (b), we have by Theorem III (b),

$$\int_{u_{*1}}^{u_1}\frac{du}{\theta(u)} = o(1), \qquad \int_{u_2}^{u_2^*}\frac{du}{\theta(u)} = o(1)$$

as $u_1 \to +\infty$, and by hypothesis

$$\lim_{u\to+\infty}\psi'(u) = \lim_{n\to+\infty}\theta'(u) = 0.$$

From these facts (5.2) follows immediately.

7. **A corollary of Theorem IV.** *If S is an L-strip with the boundary inclination $\gamma = 0$ at $u = +\infty$, for which the integrals*

$$(7.1) \qquad \int_{u_0}^{\infty}\frac{\phi_+'^2(u)}{\theta(u)}du \quad \text{and} \quad \int_{u_0}^{\infty}\frac{\phi_-'^2(u)}{\theta(u)}du \quad \text{converge,}$$

and if $z = Z(w) = X(w) + iY(w)$ maps S conformally onto the strip $|y| < \pi/2$, in such a manner that $\lim_{u\to+\infty} X(w) = +\infty$, then there exists a constant λ such that, for $w = u + iv \in S$,

$$(7.2) \quad X(w) = \lambda + \pi\int_{u_0}^{u}\frac{dt}{\theta(t)} + o(1) \quad \text{as } u \to +\infty, \text{ uniformly with respect to } v.$$

Proof. First we note that (7.1) implies the convergence of $\int_{u_0}^{\infty}(\theta'^2(u)/\theta(u))du$ and $\int_{u_0}^{\infty}(\psi'^2(u)/\theta(u))du$. Applying now Theorems III (b) and IV (b) we find that, for $w_1 = u_1 + iv_1$, $w_2 = u_2 + iv_2$ in S, $u_1 \leqq u_2$,

$$(7.3) \qquad \pi\int_{u_1}^{u_2}\frac{dt}{\theta(t)} + o(1) \leqq X(w_2) - X(w_1) \leqq \pi\int_{u_1}^{u_2}\frac{dt}{\theta(t)} + o(1),$$

where $o(1) \to 0$ as $u_1, u_2 \to +\infty$, uniformly with respect to v_1, v_2. If we set $A(w) \equiv X(w) - \pi\int_{u_0}^{u}[\theta(t)]^{-1}dt$, then (7.3) means that for every ϵ there exists an $N(\epsilon)$ such that

$$|A(w_2) - A(w_1)| < \epsilon \qquad \text{if } u_2 \geqq u_1 > N(\epsilon).$$

Hence (7.2) is true by Cauchy's convergence principle.

The condition (7.1) is very restrictive, since $\theta(u)$ might be of smaller order of magnitude than $\phi_+'^2(u)$ and $\phi_-'^2(u)$. Our next aim is, therefore, to establish, in place of Theorem III (b), a lower bound for the difference $x_2 - x_1$ in terms of the integral $\int_{u_1}^{u_2}\{(1+\psi'^2(u))/\theta(u)\}du$ and of a suitable "remainder" term. This inequality, combined with Theorem IV, will yield an asymptotic representation for $X(w)$ which will require only the existence of the integral $\int_{u_0}^{\infty}(\theta'^2(u)/\theta(u))du$ in place of (7.1).

III. THE SECOND BASIC INEQUALITY

8. **Preliminaries.** In order to derive the inequality indicated at the end of §7 we shall establish several preliminary results.

(a) Let S be an L-strip in the w-plane with the boundary inclination γ, $-\pi/2 \leqq \gamma \leqq \pi/2$, at $u = +\infty$. Moreover, let its boundary curves C_+: $v = \phi_+(u)$ and C_-: $v = \phi_-(u)$, $u \geqq u_0$, satisfy the hypothesis that $\phi_+'(u)$ and $\phi_-'(u)$ are continuous and of bounded variation in any finite interval contained in $u \geqq u_0$.

(b) Let s denote the variable arc length of C_- measured from $u = u_0$ and $w = w(s)$ the parametric representation of C_- by means of s as parameter. Let $\alpha(s)$, $-\pi/2 < \alpha(s) < \pi/2$, be the angle of inclination of the tangent to C_- at the point $w(s)$.

(c) Since the boundary inclination of S at $u = +\infty$ is γ, there exists a half-plane H_c: $u \geqq c$, such that the angle of inclination of a tangent at any point of C_+ or C_- in H_c satisfies the relation

$$|\alpha - \gamma| < \frac{\pi}{8}\cdot$$

Let C_+^* and C_-^* denote the parts of C_+ and C_-, respectively, which lie in H_c. Application of the mean value theorem shows then that any straight line with the angle of inclination β, where

$$\left|\beta - \left(\gamma + \frac{\pi}{2}\right)\right| < \frac{\pi}{8},$$

can intersect C_+^* and C_-^* in but one point each. Hence, there exists a number $\sigma > 0$ such that any normal to C_- at $w(s)$, where $s \geqq \sigma$, will intersect C_+^* and C_-^* in exactly one point each. The segment of this normal which lies in S will be denoted by Δ_s, its length by $\Delta(s)$. Evidently $\Delta(s)$ *is a continuous function for* $s \geqq \sigma$.

Denote the point $w(s)$ by A, the other end point of Δ_s (on C_+) by B and the point $u + i\phi_+(u)$ by C ($u = \Re w(s)$). Since, in the triangle ABC, the angle A approaches $|\gamma|$ and B approaches $\pi/2$ as $s \to \infty$, it follows by the law of sines that

$$\lim_{u\to+\infty} \frac{\Delta(s)}{\theta(u)} = \cos\gamma, \qquad u = \Re w(s). \tag{8.1}$$

(d) For the following two lemmas we assume that hypothesis (a) is satisfied.

LEMMA 2. *If $\alpha'(s)$ is continuous for $\sigma \leqq a \leqq s \leqq b$, and if $\Delta(s)\alpha'(s) < 1$ in the open interval $a < s < b$, then no two normals Δ_s, $a \leqq s \leqq b$, intersect in $\bar{S}$*[13].

Proof. We show first that no two normals Δ_s of any closed subarc $a < \alpha \leqq s \leqq \beta < b$ of the arc $a < s < b$, intersect. If this assertion were not true, there would exist a point A_1 $(s = a_1)$ and a point B_1 $(s = b_1)$ on C_-, $\alpha \leqq a_1 < b_1 \leqq \beta$, such that Δ_{a_1} and Δ_{b_1} would intersect at some point $D_1 \in \bar{S}$[13]. Let I_1 be the arc $a_1 \leqq s \leqq b_1$. The curvilinear triangle $A_1B_1D_1$ lies entirely within $\bar{S}$. Let $s' = (1/2)(a_1 + b_1)$. Then $\Delta_{s'}$ enters this triangle at the point $s = s'$ of C_- and will intersect either A_1D_1 or B_1D_1 or both at some point $D_2 \in \bar{S}$. If D_2 lies on A_1D_1, we let I_2 be the arc $a_2 \leqq s \leqq s'$ and denote its end points by A_2 $(s = a_2)$ and B_2 $(s = b_2)$ where $a_2 = a_1$, $b_2 = s'$. Then $\Delta_{s''}$ where $s'' = (1/2)(a_2 + b_2)$, will intersect either A_2D_2 or B_2D_2 or both at some point D_3. If D_2 does not lie on A_1D_1, let I_2 be the arc $s' \leqq s \leqq b$ where now $a_2 = s'$ and $b_2 = b_1$ and the end points of I_2 are again denoted by A_2 and B_2. Everything said about $\Delta_{s''}$ holds then for this second choice of I_2. Continuing in the manner indicated, we obtain a sequence of intervals I_n $(n = 1, 2, 3, \cdots)$ whose end points we shall denote by A_n $(s = a_n)$ and B_n $(s = b_n)$, $\alpha \leqq a_n < b_n \leqq \beta$, such that $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = c$ exist and that Δ_{a_n} and Δ_{b_n} intersect at some point $D_n \in \bar{S}$. By the law of sines, we obtain from the triangle $A_nB_nD_n$:

$$\frac{\sin D_n}{A_nB_n} = \frac{\sin A_n}{B_nD_n}.$$

Since the angle A_n approaches $\pi/2$ as $n \to \infty$ and the angle $D_n = \alpha(b_n) - \alpha(a_n) \geqq 0$, and since $B_nD_n \leqq \Delta(b_n)$ we find

$$\lim_{n\to\infty} \frac{\alpha(b_n) - \alpha(a_n)}{b_n - a_n} \geqq \lim_{n\to\infty} \frac{\sin A_n}{\Delta(b_n)} = \frac{1}{\Delta(c)}$$

or

$$\alpha'(c)\Delta(c) \geqq 1,$$

which contradicts the hypothesis.

To complete the proof of the lemma we must show that no Δ_s, $a < s < b$, intersects Δ_a or Δ_b. Suppose Δ_{s_0}, $a < s_0 < b$, intersected Δ_b at some point D_0. Then any Δ_{s_1} with $s_0 < s_1 < b$ must intersect Δ_b, since it must intersect Δ_{s_0} or Δ_b and it cannot intersect Δ_{s_0} (by the part already proved). Let B denote the point $s = b$ on C_-. The point D_1 at which Δ_{s_1} intersects Δ_b must lie on the seg-

[13] The closure of a point set M in the complex plane will be denoted by $\bar{M}$.

ment BD_0 of Δ_b (for otherwise Δ_{s_1} would meet Δ_{s_0}) and hence $BD_1 < BD_0 \leqq \Delta(b)$.

Let now $\{s_n\}$, $n=1, 2, 3, \cdots$, $s_0 < s_n < s_{n+1} < b$, denote a sequence which converges to b as $n \to \infty$, and let Δ_{s_n} intersect Δ_b at D_n. For all n, $BD_n \leqq BD_1 < \Delta(b)$. Call A_n the point $s=s_n$ on C_-. Then, by the law of sines,

$$\frac{\sin D_n}{A_nB} = \frac{\sin A_n}{BD_n} \geqq \frac{\sin A_n}{BD_1}.$$

Since the angle $A_n \to \pi/2$ as $n \to \infty$ and since $D_n = \alpha(b) - \alpha(s_n)$,

$$\lim_{n\to\infty} \frac{\alpha(b) - \alpha(s_n)}{b - s_n} \geqq \frac{1}{BD_1} > \frac{1}{\Delta(b)}, \quad \text{or} \quad \Delta(b)\alpha'(b) > 1,$$

while the continuity of $\Delta(s)\cdot\alpha'(s)$ implies that $\Delta(b)\alpha'(b) \leqq 1$. A similar argument applies to the point $s=a$.

Lemma 3. *Let Δ_a and Δ_b, $\sigma \leqq a < b$, intersect at a point C in $\bar{S}$. Suppose that any of the angles, which a chord of C_- or of C_+ through the end points*(14) *of Δ_s and $\Delta_{s'}$, $a \leqq s < s' \leqq b$, forms with Δ_s and $\Delta_{s'}$, differs from $\pi/2$ in absolute value by less than ϵ, $0 < \epsilon < \pi/8$*(15). *Then*

$$\int_a^b \frac{ds}{\Delta(s)} \leqq \frac{\alpha(b) - \alpha(a)}{\cos^3(2\epsilon)}. \tag{8.2}$$

Proof. Call the points $s=a$ and $s=b$ on C_-, A and B, respectively. Draw Δ_s, $a<s<b$, and denote its end points on C_- and C_+ by D and E, respectively. DE will intersect either AC or BC or both. Suppose it intersects BC at a point F. Call C' the end point of Δ_b on C_+. From the two triangles BDF and $C'EF$ we find by the law of sines

$$DF = BF\frac{\sin B}{\sin D}, \qquad FE = FC'\frac{\sin C'}{\sin E}.$$

Since, by hypothesis, $\sin B > \sin(\pi/2 - \epsilon) = \cos\epsilon$, $\sin C' > \cos\epsilon$, we have

$$\Delta(s) = DF + FE > BC' \cos\epsilon = \Delta(b)\cos\epsilon. \tag{8.3}$$

If Δ_s intersects Δ_a we obtain in a similar way

$$\Delta(s) \geqq \Delta(a)\cos\epsilon. \tag{8.4}$$

Now, by the mean value theorem, $\int_a^b [\Delta(s)]^{-1} ds = (b-a)/\Delta(\bar{s})$ where $a < \bar{s} < b$. Hence, by (8.3) or (8.4)

(14) If the end points of Δ_s and $\Delta_{s'}$ on C_+ coincide at a point P, then a chord through these end points on C_+ means the tangent to C_+ at P.

(15) This condition will always be satisfied if a is sufficiently large (because of the hypothesis in §8 (a)).

$$\int_a^b \frac{ds}{\Delta(s)} \leqq (b-a)\frac{1}{\Delta(a)\cos\epsilon}$$

or

$$\int_a^b \frac{ds}{\Delta(s)} \leqq (b-a)\frac{1}{\Delta(b)\cos\epsilon},$$

respectively.

From the triangle ABC we find

$$\Delta(a) \geqq AC = AB\frac{\sin B}{\sin C} \geqq AB\frac{\cos\epsilon}{\sin C}, \qquad \Delta(b) \geqq AB\frac{\cos\epsilon}{\sin C},$$

so that

$$\int_a^b \frac{ds}{\Delta(s)} \leqq \frac{\sin C}{\cos^2\epsilon}\,\frac{b-a}{AB}.$$

Observing that the hypothesis regarding the chords of C_- implies that[16]

$$b-a \leqq \frac{AB}{\cos(2\epsilon)}$$

and that the angle $C=\alpha(b)-\alpha(a)$, we obtain (8.2).

9. **"Invariant" formulation of the second basic inequality.** We prove now

THEOREM V. *Let S be an L-strip which satisfies the hypothesis of* §8 (a) *and let $\alpha(s)$, Δ_s, and $\Delta(s)$ be defined as in* §8 (b) *and* (c). *Suppose that $z=Z(w)=X(w)+iY(w)$ maps S conformally onto the strip $|y|<\pi/2$ in such a manner that $X(w)\rightarrow+\infty$ as $u\rightarrow+\infty$. Let w_1 and w_2 be points in $\bar{S}$ and Δ_{s_1} and Δ_{s_2} normals to C_- which pass through w_1 and w_2, respectively* [17]. *Let $x_1=X(w_1)$, $x_2=X(w_2)$. Then, if $s_2>s_1$,*

$$(9.1) \qquad x_2 - x_1 \geqq \pi\int_{s_1}^{s_2}\frac{ds}{\Delta(s)} - \pi(1+o(1))\int_{s=s_1}^{s_2}|d\alpha(s)| + o(1)$$

where the second integral on the right-hand side is taken in the sense of Stieltjes and $o(1)\rightarrow 0$ as $s_1\rightarrow+\infty$, uniformly in w_1 and w_2.

---

(16) The tangents to any two points of the arc AB: $a\leqq s\leqq b$, of C form an angle not exceeding 2ϵ. By the mean value theorem it follows therefore that the chord AB forms an angle of measure less than 2ϵ with the tangent to any point of the arc AB. Hence, if the arc AB is represented in the form $y=f(x)$, AB being the x-axis, A the origin and B the point $(l, 0)$, we have

$$b-a=\int_0^l (1+[f'(x)]^2)^{1/2}dx \leqq l(1+\tan^2(2\epsilon))^{1/2}=\frac{AB}{\cos(2\epsilon)}.$$

(17) If $\Re w_1$ and $\Re w_2$ are sufficiently large, there always exist normals Δ_{s_1} and Δ_{s_2} passing through w_1 and w_2, respectively.

Proof. 1. It is sufficient to prove this theorem under the assumption that $\gamma=0$ since the statement of the theorem is invariant with respect to a rotation of the coordinate system in the w-plane through the angle γ in the positive direction. Since S is an L-strip with the boundary inclination $\gamma=0$ at $u=+\infty$, there exists for every $\epsilon>0$, $\epsilon<\pi/8$, a half-plane H_c: $u\geqq c$, such that the angle of inclination α, $|\alpha|<\pi/2$, of the tangent at any point of C_+ or C_- in H_c satisfies the condition

$$|\alpha|<\epsilon/2.$$

Let an ϵ be fixed. We assume s_1 so large that *all* Δ_s *with* $s\geqq s_1$ *lie entirely in* H_c, *say* $s_1>\sigma_1$.

2. We assume first that $\alpha'(s)$ *exists and is continuous for* $s_1\leqq s\leqq s_2$. Let

$$x_2^* = \max_{w\in\Delta_{s_2}} X(w), \qquad x_{*1} = \min_{w\in\Delta_{s_1}} X(w).$$

Then we prove that

$$(9.2)\qquad x_2^* - x_{*1} \geqq \pi\int_{s_1}^{s_2}\frac{ds}{\Delta(s)} - \frac{\pi}{\cos^2(2\epsilon)}\int_{s_1}^{s_2}|\alpha'(s)|\,ds.$$

Let I_1: $a_1<s<b_1$, be a *largest open* arc of I: $s_1<s<s_2$, where

$$(9.3)\qquad \Delta(s)\alpha'(s)<1.$$

If there is no such arc we have on I

$$(9.4)\qquad \Delta(s)\alpha'(s)\geqq 1$$

and hence

$$\pi\int_{s_1}^{s_2}\frac{ds}{\Delta(s)} - \pi\int_{s_1}^{s_2}\alpha'(s)ds \leqq 0 \leqq x_2^* - x_{*1},$$

so that (9.2) is evidently true. If there are several such arcs, let I_1 be one of them. By Lemma 2, no two normals Δ_s of the closed arc $\bar I_1$: $a_1\leqq s\leqq b_1$, will intersect. If there are any normals Δ_s with $b_1<s\leqq s_2$ which intersect Δ_{b_1}, let $\Delta_{b_1'}$ be the one for which b_1' is as large as possible. Otherwise we set $b_1'=b_1$. If there are any Δ_s with $s_1\leqq s<a_1$ which intersect Δ_{a_1}, let $\Delta_{a_1'}$ be the one for which a_1' is as small as possible. Otherwise we set $a_1'=a_1$.

Call I_1' the arc $a_1'<s<b_1'$. No normal Δ_s of the arcs $(I-I_1')$ will intersect any of the Δ_s of $\bar I_1$.

If $(I-I_1')$ is void we do not proceed any further. In case $(\bar I-\bar I_1')$ is not void, either (9.4) holds for all $s\in(\bar I-\bar I_1')$, or else there exists a *largest open* arc I_2: $a_2<s<b_2$, $I_2\subset(\bar I-\bar I_1')$, where (9.3) is true. By Lemma 2 again no two normals Δ_s of $\bar I_2$ will intersect. If there are any normals Δ_s, $s\in(\bar I-\bar I_1')$, with $s>b_2$ which intersect Δ_{b_2}, let $\Delta_{b_2'}$ be the one for which b_2' is as large as possible.

Otherwise let $b_2' = b_2$. Similarly, if there are any Δ_s, $s\in(I-I_1')$, with $s<a_2$ which intersect Δ_{a_2}, let $\Delta_{a_2'}$ be the one for which a_2' is as small as possible. Otherwise let $a_2' = a_2$. Call I_2' the arc $a_2' < s < b_2'$. I_1' and I_2' have no points in common, except possibly an end point. Moreover, no normal Δ_s, $s\in\{I-(I_1'+I_2')\}$, will intersect any of the normals of I_1 and I_2, and none of the normals of I_1 will intersect any of the normals of I_2.

Continuing this construction in the manner indicated as long as possible we obtain a finite or infinite sequence of arcs I_n: $a_n<s<b_n$, in which (9.3) holds, I_n being a largest open arc of the set $(I-\sum_{k=1}^{n-1}I_k')$ for which (9.3) is true. With each I_n we obtain an arc I_n' : $a_n' < s < b_n'$ containing I_n which is defined in the following manner: If any Δ_s, $s\in(I-\sum_{k=1}^{n-1}I_k')$, with $s>b_n$ intersects Δ_{b_n}, let $\Delta_{b_n'}$ be the one for which $s=b_n'$ is a maximum, and if any Δ_s, $s\in(I-\sum_{k=1}^{n-1}I_k')$, with $s<a_n$ intersects Δ_{a_n}, let Δ_{a_n} be the one for which $s=a_n'$ is a minimum. We set $b_n'=b_n$ or $a_n'=a_n$ if there are no such Δ_s for $s>b_n$ or $s<a_n$ respectively. Then no normal Δ_s, $s\in(I-\sum_{k=1}^{n}I_k')$, intersects any of the normals Δ_s with $s\in I_k$, $(k=1, 2, 3, \cdots, n)$. This implies that the normals of I_n and I_m, $n>m$, do not intersect. For no Δ_s, $s\in(I-\sum_{k=1}^{m}I_k')$ intersects any of the Δ_s, $s\in I_k$ $(k=1, 2, \cdots, m)$, and $I_n\subset(I-\sum_{k=1}^{m}I_k')$, since $n>m$. Moreover, since, for $s\in I$, any Δ_s with $s<a_k'$ cannot intersect any of the Δ_s with $s>b_k'$ $(k=1, 2, 3, \cdots)$, it is easily seen that any two arcs I_n' and I_m', $n\neq m$ have no points in common except possibly an end point. Hence, the same is true for $(I_n'-I_n)$ and $(I_m'-I_m)$ if $m\neq n$. Let

$$A = \sum_{k=1}^{\infty} I_k, \qquad B = \bar{I} - \sum_{k=1}^{\infty} \bar{I}_k', \qquad C = \sum_{k=1}^{\infty} (\bar{I}_k' - I_k),$$

so that $\bar{I}=A+B+C$. By our construction, we have

$$\text{for } s \in A: \quad \Delta(s)\alpha'(s) < 1. \tag{9.5}$$

If B is not void, we have

$$\text{for } s \in B: \quad \Delta(s)\alpha'(s) \geqq 1. \tag{9.6}$$

If (9.6) were not true, there would be for at least one

$$s_0 \in B: \quad \Delta(s_0)\alpha'(s_0) < 1. \tag{9.7}$$

Then there exists an *open* arc $J\subset I$ containing s_0 such that (9.3) holds for $s\in J$. We show first that no arc $I_k'\subset J$. Such an arc $I_k'(a_k', b_k')$ must necessarily coincide with its subarc I_k (a_k, b_k) on which (9.3) holds, for otherwise Δ_{a_k} and $\Delta_{a_k'}$ or Δ_{b_k} and $\Delta_{b_k'}$ would intersect and that is impossible by Lemma 2 since a_k', a_k, b_k, b_k' are all points of J, where (9.3) holds. Let I_n' be the first arc in the sequence $\{I_m'\}$ (in the order in which the I_m' were constructed) which is entirely contained in J. Since s_0 is not in any I_m', I_n' lies either to the right or to the left of s_0. In either case it follows that, in choosing I_n, we did

not select the largest available interval in which (9.3) holds, since in the first case at least the arc (s_0, a_n) and in the second case the arc (b_n, s_0) should be part of I_n. Thus, no I_m' can be entirely contained in J.

Therefore, there can be in J *at most* one left end point $a_\nu' > s_0$ of some arc I_ν' and one right end point $b_\mu' < s_0$ of some I_μ' and the subarc $b_\mu' < s < a_\nu'$ of J must remain free of any points belonging to any arcs I_m' $(m = 1, 2, 3, \cdots)$. But this is impossible: The set of arcs I_m is at most denumerable and, if infinite, the lengths of the I_m approach 0 as $m \to \infty$. Since at each step of our construction a largest possible arc in which (9.3) holds is being taken for an I_m, the arc $b_\mu' < s < a_\nu'$ would have to be included at some step in the sequence $\{I_m\}$. If there is no end point a_ν' or no end point b_μ' in J, an even larger subarc of J remains free of points of any I_m', and this again is impossible. The assumption (9.7) thus leads to a contradiction, and hence (9.6) is true. We obtain therefore

$$\int_{(B)} \frac{ds}{\Delta(s)} \leqq \int_{(B)} \alpha'(s)ds = \int_{(B)} |\alpha'(s)| ds. \tag{9.8}$$

Finally, by Lemma 3, on each of the two arcs of $I_m' - I_m$, $a_m' \leqq s \leqq a_m$ and $b_m \leqq s \leqq b_m'$, we have

$$\int_{a_m'}^{a_m} \frac{ds}{\Delta(s)} \leqq \frac{1}{\cos^3(2\epsilon)} | \alpha(a_m) - \alpha(a_m') |,$$

$$\int_{b_m}^{b_m'} \frac{ds}{\Delta(s)} \leqq \frac{1}{\cos^3(2\epsilon)} | \alpha(b_m') - \alpha(b_m) |.$$

Hence,

$$\int_{(C)} \frac{ds}{\Delta(s)} \leqq \frac{1}{\cos^3(2\epsilon)} \int_{(C)} | \alpha'(s) | ds. \tag{9.9}$$

3. Consider now an arc I_n: $a_n < s < b_n$ of A. The points on a normal Δ_s are given by the equation

$$w = w(s) + ite^{i\alpha(s)}, \qquad 0 \leqq t \leqq \Delta(s). \tag{9.10}$$

The integral

$$\int_0^{\Delta(s)} | Z'(w(s) + ite^{i\alpha(s)}) | dt \equiv \int_{\Delta_s} | Z'(w) | dt \geqq \pi,$$

since it represents the length of the image of Δ_s in the z-plane. By Schwarz's inequality,

$$\frac{\pi^2}{\Delta(s)} \leqq \int_{\Delta_s} |Z'(w)|^2 dt.$$

If $s \in I_n$ and $\alpha'(s) \geqq 0$, we have, because of (9.5),

$$0 \leqq 1 - \alpha'(s)\Delta(s) \leqq 1 - \alpha'(s)t \qquad \text{for } 0 \leqq t \leqq \Delta(s);$$

and if $\alpha'(s) < 0$, we have

$$1 \leqq 1 - \alpha'(s)t \qquad \text{for } 0 \leqq t \leqq \Delta(s).$$

Hence, in either case

$$-\pi^2 \int_{a_n}^{b_n} |\alpha'(s)|\, ds + \pi^2 \int_{a_n}^{b_n} \frac{ds}{\Delta(s)} \leqq \int_{a_n}^{b_n} ds \int_0^{\Delta(s)} |Z'(w)|^2 (1 - \alpha'(s)t)\, dt.$$

By what was said in part 2, no two normals Δ_s, $a_n \leqq s \leqq b_n$, will intersect, and the transformation (9.10) maps therefore the region $\{0 < t < \Delta(s), a_n < s < b_n\}$, of an (s, t)-plane in a one-to-one manner onto the (limited) subregion T_n of S which is bounded by C_-, C_+, Δ_{a_n} and Δ_{b_n}. The Jacobian of this transformation is $(1 - \alpha'(s)t)$. Hence, the last double integral may be written as

$$\iint_{(T_n)} |Z'(u + iv)|^2 du dv.$$

Summation over n gives

$$(9.11) \qquad -\pi^2 \int_{(A)} |\alpha'(s)|\, ds + \pi^2 \int_{(A)} \frac{ds}{\Delta(s)} \leqq \sum_{n=1}^{\infty} \iint_{(T_n)} |Z'(u + iv)|^2 du dv.$$

Since, again by the discussion of part 2, the normals Δ_s of any I_n do not intersect those of any other I_m $(n \neq m)$, *the regions* T_n *for* $n = 1, 2, 3, \cdots$ *do not overlap*. Each T_n is mapped by $z = Z(w)$ onto a subregion of the strip $|y| < \pi/2$ whose area is given by $\iint_{(T_n)} |Z'(u+iv)|^2 du dv$. Since the T_n do not overlap, their images in the z-plane do not overlap either, and the total area which these images cover is therefore less than or equal to $\pi(x_2^* - x_{*1})$. Hence, from (9.11)

$$(9.12) \qquad -\pi \int_{(A)} |\alpha'(s)|\, ds + \pi \int_{(A)} \frac{ds}{\Delta(s)} \leqq x_2^* - x_{*1}.$$

Combining (9.12) with (9.8) and (9.9) we find (9.2).

4. In order to prove now the general case of the theorem (where $\alpha'(s)$ is not necessarily continuous) we approximate the arc β: $w = w(s)$, $s_1 \leqq s \leqq s_2$ of C_- by a sequence of arcs β_n: $w = w_n(s)$, $s_1 \leqq s \leqq s_2$ $(n = 1, 2, \cdots)$, with the following properties([18]):

(i) $w_n(s)$, $w_n'(s)$, $w_n''(s)$ are continuous for $s_1 \leqq s \leqq s_2$, $w_n(s_k) = w(s_k)$, $w_n'(s_k) = w'(s_k)$, $(k = 1, 2)$, and β_n does not intersect any part of the boundary of S except β.

([18]) The functions $w_n(s)$ may be found as follows: Let $\alpha(s; h) = (1/h)\int_s^{s+h} \alpha(t)dt$, $s \geqq s_1$, $h > 0$. Then $\alpha(s; h) \to \alpha(s)$ as $h \to 0$, uniformly for $s_1 \leqq s \leqq s_2$. Let $\alpha(s) = \alpha_1(s) - \alpha_2(s)$, where $\alpha_1(s)$, $\alpha_2(s)$ are continuous, non-decreasing functions. The total variation of $\alpha(s; h)$ in $s_1 \leqq s \leqq s_2$ is

(ii) $w_n(s) \to w(s)$ and $w_n'(s) \to w'(s)$ as $n \to \infty$, uniformly for $s_1 \leqq s \leqq s_2$. This condition implies that if $\sigma = \sigma_n(s)$ is the variable arc length of β_n, measured from $s = s_1$, and if $\alpha_n(s)$ is the angle $(|\alpha_n(s)| < \pi/2)$ which the tangent to β_n at $w_n(s)$ forms with the positive real axis, then

$$\sigma_n(s) \to s - s_1, \qquad \alpha_n(s) \to \alpha(s) \qquad \text{as } n \to \infty, \tag{9.13}$$

uniformly for all $s_1 \leqq s \leqq s_2$.

(iii) The total variation

$$\int_{s=s_1}^{s_2} |\, d\alpha_n(s)\,| \to \int_{s=s_1}^{s_2} |\, d\alpha(s)\,| \qquad \text{as } n \to \infty. \tag{9.14}$$

Denote by S_n the region obtained from S when the arc β is replaced by β_n. For sufficiently large n and $s \geqq s_1$, $\Delta_n(s)$ can be defined for S_n as $\Delta(s)$ is defined for S (see part 1 and §8 (c)). Condition (ii) implies that, as $n \to \infty$,

$$\Delta_n(s) \to \Delta(s) \qquad \textit{uniformly for } s_1 \leqq s \leqq s_2. \tag{9.15}$$

By a well known theorem, it follows from (9.13) and (9.15) that

$$\int_{s=s_1}^{s_2} \frac{d\sigma_n(s)}{\Delta_n(s)} \to \int_{s_1}^{s_2} \frac{ds}{\Delta(s)}. \tag{9.16}$$

Finally let S_n be mapped onto $|y| < \pi/2$ so that $u = +\infty$ and $x = +\infty$ correspond to each other and that a point w_0 which is in S and in all S_n is carried

---

$$\int_{s=s_1}^{s_2} |\, d\alpha(s;h)| \leqq \frac{1}{h}\int_{s=s_1}^{s_2} d\left[\int_s^{s+h} \alpha_1(t)dt\right] + \frac{1}{h}\int_{s=s_1}^{s_2} d\left[\int_s^{s+h} \alpha_2(t)dt\right].$$

Since

$$\frac{1}{h}\int_{s=s_1}^{s_2} d\left[\int_s^{s+h} \alpha_i(t)dt\right] = \frac{1}{h}\int_{s_1}^{s_2}[\alpha_i(s+h) - \alpha_i(s)]ds = \frac{1}{h}\int_{s_1}^{s_2} ds \int_{t=s}^{s+h} d\alpha_i(t)$$
$$\leqq \frac{1}{h}\int_{s=s_1}^{s_2+h} d\alpha_i(t)\int_{t-h}^{t} ds = \int_{s=s_1}^{s_2+h} d\alpha_i(t),$$

we have

$$\int_{s=s_1}^{s_2} |\, d\alpha(s;h)| \leqq \int_{t=s_1}^{s_2+h} d\alpha_1(t) + \int_{t=s_1}^{s_2+h} d\alpha_2(t).$$

Thus, for $h \leqq h_0$, the total variation of $\alpha(s;h)$ in $s_1 \leqq s \leqq s_2$ is uniformly bounded, and there exists, therefore, by a theorem of E. Helly, a sequence $h_n \to 0$ as $n \to \infty$, such that, for $\bar{\alpha}_n(s) \equiv \alpha(s;h_n)$:

$$\int_{s_1}^{s_2} |\, \bar{\alpha}_n'(s)|\, ds = \int_{s=s_1}^{s_2} |\, d\bar{\alpha}_n(s)| \to \int_{s=s_1}^{s_2} |\, d\alpha(s)| \qquad \text{as } n \to \infty.$$

Let $w_1^* = w(s_1)$, $w_2^* = w(s_2)$. Then we set

$$w_n(s) = w_1^* + \frac{w_2^* - w_1^*}{g_n}\int_{s_1}^{s} e^{i\bar{\alpha}_n(t)}dt + (a_n s + b_n)(s - s_1)(s - s_2),$$

where $g_n = \int_{s_1}^{s_2} e^{i\bar{\alpha}_n(t)}dt$ and a_n and b_n are so determined that $w_n'(s_1) = w'(s_1)$, $w_n'(s_2) = w'(s_2)$. It is easily seen that $\lim_{n\to+\infty} g_n = w_2^* - w_1^*$, $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = 0$ and that the $w_n(s)$ satisfy the three conditions (i), (ii), (iii) of the text.

into $Z(w_0)$. Let $x_{*1}^{(n)}$ and $x_2^{*(n)}$ be defined for S_n as x_{*1} and x_2^* are for S. Then[19]

$$(9.17)\qquad \lim_{n\to\infty} x_{*1}^{(n)} = x_{*1} \quad\text{and}\quad \lim_{n\to\infty} x_2^{*(n)} = x_2^*.$$

For sufficiently large n, S_n will satisfy the condition stated at the end of part 1 of this proof, since S satisfies it. Moreover, since β_n has continuous curvature, we have by (9.2)

$$-\frac{\pi}{\cos^3(2\epsilon)}\int_{s=s_1}^{s_2} |\,d\alpha_n(s)\,| + \pi\int_{s_1}^{s_2}\frac{d\sigma_n(s)}{\Delta_n(s)} \leqq x_2^{*(n)} - x_{*1}^{(n)}.$$

Letting $n\to\infty$ we find because of (9.14), (9.16), and (9.17),

$$-\frac{\pi}{\cos^3(2\epsilon)}\int_{s=s_1}^{s_2} |\,d\alpha(s)\,| + \pi\int_{s_1}^{s_2}\frac{ds}{\Delta(s)} \leqq x_2^* - x_{*1}.$$

Finally, to prove (9.1) we need only observe that by Lemma 1 $x_2^*-x_2\to 0$ and $x_1-x_{*1}\to 0$ as $s_1\to\infty$ and that ϵ may be taken arbitrarily small, provided s_1 is chosen sufficiently large. This completes the proof of Theorem V.

10. **A definition and further lemmas.** Theorem V will enable us to obtain a lower bound for the difference x_2-x_1 of the type mentioned in §7. We introduce first the following definition.

DEFINITION. *Let S be an L-strip in the w-plane with the boundary inclination γ, $|\gamma|<\pi/2$, at $u=+\infty$. Let $v=\phi_+(u)$ and $v=\phi_-(u)$ represent the boundary curves C_+ and C_- of S, respectively. We shall say that S is a strip with finite boundary turning at $u=+\infty$ if $\phi_+'(u)$ and $\phi_-'(u)$ are continuous for all sufficiently large u, say $u\geqq u_1$, and if the integrals*

$$\int_{u=u_1}^{\infty} |\,d\phi_+'(u)\,|, \qquad \int_{u=u_1}^{\infty} |\,d\phi_-'(u)\,|$$

converge.

Our object will be accomplished by the following lemma.

LEMMA 4. *Let S be an L-strip in the w-plane with the boundary inclination γ, $|\gamma|<\pi/2$, and with finite boundary turning at $u=+\infty$. If $\Delta(s)$ and $w(s)$ are defined as in §8, then for $u=\Re w(s)$ and $s_1<s_2$ ($u_1=\Re w(s_1)$, $u_2=\Re w(s_2)$)*

$$(10.1)\qquad \int_{s_1}^{s_2}\frac{ds}{\Delta(s)} = \int_{u_1}^{u_2}\frac{1+\psi'^2(u)}{\theta(u)}\,du - \frac{1}{4}\int_{u_1}^{u_2}\frac{\theta'^2(u)}{\theta(u)}\,du + o(1) \quad\text{as } s_1\to\infty.$$

Proof. Denote the point $w(s)$ on C_- by A, the other end point of Δ_s by B and the point $u+i\phi_+(u)$, where $u=\Re w(s)$, by C. Suppose, for the present, that B lies in the half-plane to the left of AC. In the triangle ABC the angle

(19) This follows from the theorem of Radó used in the proof of (6.9); see (12).

$A = \alpha = \alpha(s)$. To find the angle C, observe that by the mean value theorem there exists a point $w^* = u^* + i\phi_+(u^*)$ on the arc BC of C_+ such that the tangent to C_+ at w^* is parallel to the chord BC. If the angle of inclination of this tangent is β^*, $|\beta^*| < \pi/2$, then the angle $C = \pi/2 - \beta^*$. Let β, $|\beta| < \pi/2$, denote the angle of inclination of the tangent to C_+ at C. By the law of sines we obtain from the triangle ABC:

$$(10.2)\quad \frac{AC}{AB} = \frac{\theta(u)}{\Delta(s)} = \frac{\sin(A + C)}{\sin C} = \frac{\cos(\beta^* - \alpha)}{\cos \beta^*} = \frac{\cos[(\beta - \alpha) + (\beta^* - \beta)]}{\cos[\beta + (\beta^* - \beta)]}.$$

Since α, β, β^* approach γ as $s \to \infty$ and since $|\gamma| < \pi/2$, we may write (10.2) in the form

$$\frac{\theta(u)}{\Delta(s)} = \frac{\cos(\beta - \alpha)}{\cos \beta}(1 + O(|\beta^* - \beta|)) \qquad \text{as } s \to \infty.$$

Now

$$|\beta^* - \beta| \leqq |\tan \beta^* - \tan \beta| \leqq \left| \int_{t=u}^{u^*} |d\phi_+'(t)| \right|.$$

Moreover,

$$|u - u^*| \leqq BC = \frac{\sin A}{\sin B}\theta(u) < \theta(u)$$

for sufficiently large u, since $B \to \pi/2$ and $A \to |\gamma|$ as $u \to +\infty$. Hence

$$(10.3)\qquad |\beta^* - \beta| \leqq \Phi_+(u) \equiv \int_{t=u-\theta(u)}^{u+\theta(u)} |d\phi_+'(t)|.$$

The quotient

$$\frac{\cos(\beta - \alpha)}{\cos \beta \cos \alpha} = 1 + \tan \alpha \tan \beta = 1 + \phi_+'(u)\phi_-'(u) = 1 + \psi'^2(u) - \left[\frac{\theta'(u)}{2}\right]^2,$$

so that

$$\frac{1}{\Delta(s)} = \frac{\cos \alpha}{\theta(u)}(1 + \psi'^2(u)) - \frac{\cos \alpha}{\theta(u)}\left[\frac{\theta'(u)}{2}\right]^2 + O\left(\frac{\Phi_+(u)}{\theta(u)}\cos \alpha\right)$$

as $u \to +\infty$. This relation is also easily verified if B lies to the right of AC. Observing now that $\cos \alpha \cdot ds = du$, we obtain

$$\int_{s_1}^{s_2} \frac{ds}{\Delta(s)} = \int_{u_1}^{u_2} \frac{1 + \psi'^2(u)}{\theta(u)} du - \frac{1}{4}\int_{u_1}^{u_2} \frac{\theta'^2(u)}{\theta(u)} du + O\left\{ \int_{u_1}^{u_2} \frac{\Phi_+(u)}{\theta(u)} du \right\}.$$

The result (10.1) follows now immediately from

LEMMA 5. *Under the hypotheses of Lemma 4 the integral*

$$(10.4)\qquad \int^{\infty} \frac{\Phi_+(u)}{\theta(u)}\,du \quad \textit{converges},\quad \Phi_+(u) \equiv \int_{t=u-\theta(u)}^{u+\theta(u)} |\,d\phi_+'(t)\,|.$$

Proof. We have for $b<c$:

$$I \equiv \int_b^c \frac{\Phi_+(u)}{\theta(u)}\,du = \int_b^c \frac{1}{\theta(u)}\left[\int_{t=u}^{u+\theta(u)} |\,d\phi_+'(t)\,|\right]du$$
$$+ \int_b^c \frac{1}{\theta(u)}\left[\int_{t=u-\theta(u)}^{u} |\,d\phi_+'(t)\,|\right]du = I_1 + I_2.$$

Interchanging of the order of integration in I_1 gives

$$I_1 = \int_{t=b}^{c} |\,d\phi_+'(t)\,| \int_{g(t)}^{t} \frac{du}{\theta(u)} + \int_{t=c}^{c+\theta(c)} |\,d\phi_+'(t)\,| \int_{g(t)}^{c} \frac{du}{\theta(u)}$$
$$\leqq \int_{t=b}^{c+\theta(c)} |\,d\phi_+'(t)\,| \int_{g(t)}^{t} \frac{du}{\theta(u)},$$

where $u=g(t)$ is the inverse function([20]) of $t=u+\theta(u)$ for $b+\theta(b)\leqq t\leqq c+\theta(c)$, $u=g(t)\equiv b$ for $b\leqq t\leqq b+\theta(b)$. Now $0\leqq t-g(t)\leqq\theta(g(t))$ for $b\leqq t\leqq c+\theta(c)$. Hence, by Lemma 7 (which is proved in §15),

$$\theta(u) \geqq \tfrac{1}{2}\theta(g(t)) \qquad\qquad \text{for } g(t) \leqq u \leqq t,$$

provided only b is sufficiently large. Thus

$$\int_{g(t)}^{t} du/\theta(u) \leqq 2\left[(t-g(t))/\theta(g(t))\right] \leqq 2,$$

and hence

$$I_1 \leqq 2\int_{t=b}^{c+\theta(c)} |\,d\phi_+'(t)\,|.$$

Similarly,

$$I_2 \leqq 2\int_{t=b-\phi(b)}^{c} |\,d\phi_+'(t)\,|,$$

if b is sufficiently large, so that

$$I \leqq 4\int_{t=b-\theta(b)}^{c+\theta(c)} |\,d\phi_+'(t)\,|.$$

Hence (10.4) follows from the hypothesis that S has finite boundary turning at $u=+\infty$.

11. **The second basic inequality.** We are now in the position to prove our second basic inequality.

([20]) The inverse function of $t=u+\theta(u)$ exists if b is sufficiently large, since $dt/du=1+\theta'(u)$, and therefore approaches 1 as $u\to+\infty$.

THEOREM VI. *Let S be an L-strip in the w-plane with the boundary inclination $\gamma=0$ and with finite boundary turning at $u=+\infty$. Suppose that $z=Z(w)=X(w)+iY(w)$ maps S conformally onto the strip $|y|<\pi/2$ and that $\lim_{u\to+\infty} X(w)=+\infty$. Let $w_1=u_1+iv_1$, $w_2=u_2+iv_2$, $u_1<u_2$, be two points in $\bar{S}$ and let $x_1=X(w_1)$, $x_2=X(w_2)$. Then*

$$x_2 - x_1 \geqq \pi \int_{u_1}^{u_2} \frac{1+\psi'^2(u)}{\theta(u)}\,du - \frac{\pi}{4}\int_{u_1}^{u_2} \frac{\theta'^2(u)}{\theta(u)}\,du + o(1) \tag{11.1}$$

as u_1 and hence u_2 approach $+\infty$, uniformly in v_1 and v_2.

Proof. Let $w(s)$ be defined as in §8 (b), and let $\Re w(s_k)=u_k$, $X(u_k+i\phi_-(u_k)) =x_k'$ $(k=1, 2)$. By Theorem V we have, since $\int_{t=u_1}^{\infty}|d\phi_-'(t)|<\infty$,

$$x_2' - x_1' \geqq \pi \int_{s_1}^{s_2} \frac{ds}{\Delta(s)} + o(1) \text{ as } s_1 \text{ and hence } s_2 \to \infty. \tag{11.2}$$

By Lemma 1, $\lim_{u_2\to+\infty}(x_2-x_2')=0$ and $\lim_{u_1\to+\infty}(x_1-x_1')=0$ so that the difference $x_2'-x_1'$ in (11.2) may be replaced by $x_2-x_1+o(1)$. The result (11.1) follows now immediately from Lemma 4.

IV. A DISTORTION THEOREM

12. **A lemma.** *If in Theorems* IV *and* VI *it is assumed that the integral*

$$\int^{\infty} \frac{\theta'^2(u)}{\theta(u)}\,du \quad \textit{converges,}$$

then their combination yields an asymptotic expression for x_2-x_1,

$$x_2 - x_1 = \int_{u_1}^{u_2} \frac{1+\psi'^2(u)}{\theta(u)}\,du + o(1) \textit{ as } u_1 \textit{ and } u_2 \to +\infty. \tag{12.1}$$

One of the hypotheses under which this formula is obtained is that the boundary inclination of S at $u=+\infty$ is $\gamma=0$. It is easy to modify it so as to obtain a result which holds for any γ, $|\gamma|<\pi/2$. For this purpose we prove first the following lemma.

LEMMA 6. *Let S be an L-strip in the w-plane $(w=u+iv)$ with the boundary inclination γ, $|\gamma|<\pi/2$, and with finite boundary turning at $u=+\infty$. Let $v=\phi_+(u)$ and $v=\phi_-(u)$ represent its boundary curves C_+ and C_-, respectively. Suppose that a new set of coordinates $\bar{u}$, $\bar{v}$ $(\bar{u}+i\bar{v}=\bar{w})$ is introduced by means of the rotation $\bar{w}=we^{-i\gamma}$ and that, for sufficiently large $\bar{u}$, C_+ and C_- are represented in the new coordinate system by $\bar{v}=\bar{\phi}_+(\bar{u})$ and $\bar{v}=\bar{\phi}_-(\bar{u})$, respectively. Set*

$$\theta(u) = \phi_+(u) - \phi_-(u), \qquad \bar{\theta}(\bar{u}) = \bar{\phi}_+(\bar{u}) - \bar{\phi}_-(\bar{u}).$$

If u and $\bar{u}$ are connected by the relation

$$(12.2)\qquad \bar{u} = u\cos\gamma + \phi_-(u)\sin\gamma,$$

then, for sufficiently large u,

$$(12.3)\qquad |\bar{\theta}'(\bar{u})| \leqq 2\{|\theta'(u)| + \Phi_+(u)\}, \qquad \Phi_+(u) \equiv \int_{t=u-\theta(u)}^{u+\theta(u)} |d\phi_+'(t)|.$$

Thus, the convergence of the integral

$$\int^{\infty} \frac{\theta'^2(u)}{\theta(u)}\,du$$

implies that of

$$\int^{\infty} \frac{\bar{\theta}'^2(\bar{u})}{\bar{\theta}(\bar{u})}\,d\bar{u}$$

(*by* (8.1) *and Lemma* 5).

Proof. Let u and $\bar{u}$ satisfy (12.2). We denote the points $\bar{u}+i\bar{\phi}_-(\bar{u})$ and $\bar{u}+i\bar{\phi}_+(\bar{u})$ of the $\bar{w}$-plane by A and B, respectively. The coordinates of A and B in the w-plane are then $u+i\phi_-(u)$ and $u_1+i\phi_+(u_1)$, respectively (where u_1 is determined by the relation $e^{-i\gamma}(u_1+i\phi_+(u_1)) = \bar{u}+i\bar{\phi}_+(\bar{u})$). Finally, we call C the point $u+i\phi_+(u)$ of the w-plane. In the triangle ABC, $AB=\bar{\theta}(\bar{u})$, $AC=\theta(u)$ and the angle $A=|\gamma|$.

Let $\alpha_+(u)$ and $\alpha_-(u)$ denote in the w-plane the angles of inclination of the tangents to C_+ and to C_-, respectively, at a point with the abscissa u. Then

$$\phi_-'(u) = \tan\alpha_-(u), \qquad \phi_+'(u) = \tan\alpha_+(u);$$

$$\bar{\phi}_-'(\bar{u}) = \tan\{\alpha_-(u)-\gamma\}, \qquad \bar{\phi}_+'(\bar{u}) = \tan\{\alpha_+(u_1)-\gamma\}.$$

Hence,

$$\bar{\theta}'(\bar{u}) = \tan\{\alpha_+(u_1)-\alpha_-(u)\}\cdot\{1+\bar{\phi}_-'(\bar{u})\bar{\phi}_+'(\bar{u})\}.$$

Since $\bar{\phi}_-'(\bar{u})$ and $\bar{\phi}_+'(\bar{u})$ approach 0 as $\bar{u}\to+\infty$, we may write:

$$(12.4)\qquad |\bar{\theta}'(\bar{u})| \leqq \{|\theta'(u)| + |\phi_+'(u_1)-\phi_+'(u)|\}(1+o(1))$$

as $u\to+\infty$. Now

$$|u_1 - u| \leqq BC = \frac{\sin A}{\sin B}\,\theta(u) \leqq \theta(u)$$

for sufficiently large u, since the angle $B\to\pi/2$ and $A=|\gamma|<\pi/2$. Hence, $|\phi_+'(u_1)-\phi_+'(u)| \leqq \left|\int_{t=u}^{u_1}|d\phi_+'(t)|\right| \leqq \Phi_+(u)$, and (12.3) follows from (12.4).

13. **First form of the distortion theorem.** We prove now

THEOREM VII. *Let S be an L-strip with the boundary inclination* γ, $|\gamma|<\pi/2$ *and with finite boundary turning at* $u=+\infty$. *Moreover let the integral*

$$\int^{\infty} \frac{\theta'^2(u)}{\theta(u)} du \quad \textit{be convergent}(^{21}). \tag{13.1}$$

Suppose that $z=Z(w)=X(w)+iY(w)$ *maps* S *conformally onto the strip* $|y|<\pi/2$ *in such a manner that* $\lim_{u\to+\infty} X(w)=+\infty$. *Let* Δ_s, $\Delta(s)$ *and* $w(s)$ *be defined as in* §8. *Then, if* $s_1<s_2$, $w_1\in\Delta_{s_1}$, $w_2\in\Delta_{s_2}$,

$$X(w_2) - X(w_1) = \pi \int_{s_1}^{s_2} \frac{ds}{\Delta(s)} + o(1) \qquad \textit{as } s_1 \textit{ and } s_2 \to \infty, \tag{13.2}$$

uniformly with respect to w_1 *and* w_2.

Proof. As in Lemma 6, let there be introduced a new set of axes $(\bar{u}, \bar{v})$, $\bar{u}+i\bar{v}=\bar{w}$, by means of the rotation $\bar{w}=e^{-i\gamma}w$. Let $\bar{\phi}_+(u)$, $\bar{\phi}_-(u)$, $\bar{\theta}(\bar{u})$ be defined as in Lemma 6 and let u and $\bar{u}$ be two numbers related by equation (12.2). Then, by Lemma 6, (13.1) implies that

$$\int^{\infty} \frac{\bar{\theta}'^2(\bar{u})}{\bar{\theta}(\bar{u})} d\bar{u} \tag{13.3}$$

converges. Let $\bar{u}_1$ and $\bar{u}_2$ be the abscissas of the points $w(s_1)$ and $w(s_2)$ (on C_-) in the $(\bar{u}, \bar{v})$-system, and let $x_1' = X(w(s_1))$, $x_2' = X(w(s_2))$. If $\bar{\psi}(\bar{u}) = \frac{1}{2}[\bar{\phi}_+(\bar{u})+\bar{\phi}_-(\bar{u})]$, we have by Theorem IV,

$$x_2' - x_1' \leqq \pi \int_{\bar{u}_1}^{\bar{u}_2} \frac{1 + \bar{\psi}'^2(\bar{u})}{\bar{\theta}(\bar{u})} d\bar{u} + \frac{\pi}{12}\int_{\bar{u}_1}^{\bar{u}_2} \frac{\bar{\theta}'^2(\bar{u})}{\bar{\theta}(\bar{u})} d\bar{u} + o(1),$$

as s_1 and $s_2\to\infty$. By Lemma 4 and by (13.3) we obtain

$$x_2' - x_1' \leqq \int_{s_1}^{s_2} \frac{ds}{\Delta(s)} + o(1) \qquad \text{as } s_1, s_2 \to \infty. \tag{13.4}$$

Finally, application of Lemma 1 shows that $\lim_{s_1\to\infty} (x_1-x_1') = \lim_{s_1\to\infty} (x_2-x_2') = 0$ $(x_k=X(w_k))$, so that the difference $x_2' - x_1'$ in (13.4) may be replaced by x_2-x_1. Combining this result with Theorem V, we obtain (13.2).

14. **Second form of the distortion theorem.** For some applications the following form of Theorem VII will be more convenient.

THEOREM VIII. *Let* S *be an* L*-strip satisfying all the hypotheses of Theorem* VII, *and let* $z=Z(w)$ *be defined as in that theorem. Then for* $w_1=u_1+iv_1$ *and* $w_2=u_2+iv_2$ *in* $\bar{S}$, $u_1<u_2$,

---

(21) Condition (13.1) is automatically satisfied for L-strips with finite boundary turning at $u=+\infty$ for which $0<c_1\leqq\theta(u)\leqq c_2$ (c_1, c_2 constants). For

$$\int_a^b \frac{\theta'^2(u)}{\theta(u)} du = \theta'(u) \log \theta(u) \Big|_a^b - \int_{a-0}^{b} \log \theta(u) d\theta'(u).$$

Since $|\log \theta(u)|$ is bounded and $\int_{-\infty}^{\infty} |d\theta'(u)|$ exists, (13.1) follows.

$$X(w_2) - X(w_1) = \pi \int_{u_1}^{u_2} \frac{1 + \psi'^2(u)}{\theta(u)} du$$
$$+ \pi \left[\frac{v_2 - \psi(u_2)}{\theta(u_2)} - \frac{v_1 - \psi(u_1)}{\theta(u_1)} \right] \tan \gamma + o(1),$$

as u_1 and $u_2 \to +\infty$, uniformly with respect to v_1 and v_2.

Proof. Let Δ_s, $\Delta(s)$ and $w(s)$ be defined as in §8. Let Δ_{s_1} and Δ_{s_2} be two normals of C_- which pass through w_1 and w_2, respectively. By Theorem VII,

$$X(w_2) - X(w_1) = \pi \int_{s_1}^{s_2} \frac{ds}{\Delta(s)} + o(1) \qquad \text{as } s_1, s_2 \to \infty.$$

Let $\Re w(s_1) = u_1'$, $\Re w(s_2) = u_2'$. By Lemma 4 and (13.1) this equals

$$\pi \int_{u_1'}^{u_2'} \frac{1 + \psi'^2(u)}{\theta(u)} du + o(1), \qquad \text{as } u_1, u_2 \to +\infty,$$

or

$$X(w_2) - X(w_1) = \pi \int_{u_1'}^{u_1} + \pi \int_{u_1}^{u_2} + \pi \int_{u_2}^{u_2'} + o(1). \tag{14.1}$$

We estimate the first and third of the integrals in (14.1). Assume, for the present, that $\gamma > 0$. Denote the points $u_1 + i\phi_-(u_1)$, $w(s_1)$ and w_1 by A, B, and C, respectively. In the triangle ABC the side $AC = v_1 - \phi_-(u_1)$, the angle $C \to \gamma$ and the angle $B \to \pi/2$ as $u_1 \to +\infty$. *These limits, as well as all following in this proof (taken as $u_1 \to +\infty$) exist uniformly with respect to the position of w_1 on θ_{u_1}.* Now

$$u_1' - u_1 = BC \sin(\gamma + \epsilon_1), \qquad \lim_{u_1 \to +\infty} \epsilon_1 = 0,$$

$$BC = AC \frac{\sin A}{\sin B} = AC \frac{\cos(\gamma + \epsilon_2)}{\sin B}, \qquad \lim_{u_1 \to +\infty} \epsilon_2 = 0,$$

and hence

$$u_1' - u_1 = [v_1 - \phi_-(u_1)] \sin(\gamma + \epsilon_1) \cos(\gamma + \epsilon_2) \frac{1}{\sin B}. \tag{14.2}$$

This result is also easily verified when $\gamma \leq 0$.

Now, by the law of the mean

$$\int_{u_1}^{u_1'} \frac{1 + \psi'^2(u)}{\theta(u)} du = (u_1' - u_1) \frac{1 + \psi'^2(u^*)}{\theta(u^*)} \qquad (u^* \text{ between } u_1 \text{ and } u_1').$$

Since, by (14.2), $|u_1' - u_1| \leq \theta(u_1)$ for sufficiently large u_1, it follows from Lemma 7 (which is proved in §15) that $(\theta(u^*)/\theta(u_1)) \to 1$ as $u_1 \to +\infty$. Furthermore, as $u_1 \to +\infty$, $1 + \psi'^2(u^*) \to 1/\cos^2 \gamma$. Hence

$$\int_{u_1}^{u'} \frac{1+\psi'^2(u)}{\theta(u)}\,du = \frac{v_1 - \phi_-(u_1)}{\theta(u_1)} \tan\gamma + o(1)$$

$$= \left[\frac{v_1 - \psi(u_1)}{\theta(u_1)} + \frac{1}{2}\right] \tan\gamma + o(1)$$

as $u_1 \to +\infty$. An analogous result is obtained for the third integral in (14.1). Substitution of these expressions into (14.1) proves the theorem.

COROLLARY OF THEOREM VIII. *Under the hypotheses of the theorem, for* $w = u + iv \in \bar{S}$,

$$\lim_{u\to+\infty} \left\{ X(w) - \left[\pi \int_{u_0}^{u} \frac{1+\psi'^2(t)}{\theta(t)}\,dt + \pi \frac{v-\psi(u)}{\theta(u)} \tan\gamma\right]\right\} = \lambda \tag{14.3}$$

exists uniformly in v and is finite.

For, if the difference within the braces in (14.3) is denoted by $A(w)$, Theorem VIII states that for any given $\epsilon > 0$ there exists an $N(\epsilon)$ such that for all w_1 and w_2 in S for which $\Re w_2 > \Re w_1 > N(\epsilon)$:

$$|A(w_2) - A(w_1)| < \epsilon.$$

THEOREM IX. *Let S be an L-strip satisfying the hypotheses of Theorem* VII *and let* $z = Z(w)$ *be defined as in that theorem. Then, for* $w = u + iv \in \bar{S}$,

$$Z(w) = \lambda + \pi \int_{u_0}^{u} \frac{1+\psi'^2(t)}{\theta(t)}\,dt + \pi \frac{v-\psi(u)}{\theta(u)} \tan\gamma + i\pi \frac{v-\psi(u)}{\theta(u)} + o(1),$$

as $u \to +\infty$, *uniformly in v.* λ *is a real constant.*

This result follows immediately by combining (14.3) with the result (18.1) of Corollary 1 of Theorem X (which is proved in §18).

V. ASYMPTOTIC BEHAVIOR OF THE MAPPING FUNCTION OF AN L-STRIP AND OF ITS DERIVATIVE

15. **Preliminary remarks and lemmas.** We shall establish now asymptotic expressions for $Z(w)$ and $Z'(w)$ under the mere assumption that S is an L-strip. These results will be less sharp than those of Part IV. While in Part IV (Theorem IX) we obtained, under more restrictive assumptions, an expression $f(w)$ such that the *difference* $Z(w) - f(w)$ approaches a finite limit, as $u \to +\infty$, we shall find here expressions for $Z(w)$ and $Z'(w)$ which represent these functions merely in the sense of *asymptotic equivalence* (that is, the *quotient* of the function in question and its asymptotic expression approaches 1 as $u \to +\infty$). None of the results of Parts III and IV will be used here.

Throughout this part we assume S to be an L-strip in the w-plane with the boundary inclination γ, $|\gamma| < \pi/2$ at $u = +\infty$, and $Z(w) = X(w) + iY(w)$

a function which maps S conformally onto the strip $|y| < \pi/2$ in such a manner that $\lim_{u\to+\infty} X(w) = +\infty$. The inverse of $z = Z(w)$ will be denoted by $w = W(z) = U(z) + iV(z)$.

We shall make use of the following simple lemma.

LEMMA 7. *Let, for* $u_0 < u_1 < u_2$,

$$\theta^*(u_1, u_2) = \max_{u_1 \leqq u \leqq u_2} \theta(u), \qquad \theta_*(u_1, u_2) = \min_{u_1 \leqq u \leqq u_2} \theta(u)$$

and let

$$u_2 - u_1 \leqq k\theta^*(u_1, u_2), \qquad k \text{ a constant.} \tag{15.1}$$

Then, uniformly for all u_2,

$$\lim_{u_1 \to +\infty} \frac{\theta_*(u_1, u_2)}{\theta^*(u_1, u_2)} = 1.$$

Proof. Let $\theta^*(u_1, u_2) = \theta(b)$ and $\theta_*(u_1, u_2) = \theta(c)$, $u_1 \leqq b$, $c \leqq u_2$. If u_1 is sufficiently large, we have

$$|\theta(b) - \theta(c)| = \left| \int_c^b \theta'(u) du \right| \leqq \int_{u_1}^{u_2} |\theta'(u)| du \leqq (u_2 - u_1) \sup_{u_1 \leqq u \leqq u_2} |\theta'(u)|,$$

(where $\sup_{u_1 \leqq u \leqq u_2} |\theta'(u)|$ denotes the least upper bound of $|\theta'(u)|$ in $u_1 \leqq u \leqq u_2$). Hence by use of (15.1):

$$0 \leqq 1 - \frac{\theta_*(u_1, u_2)}{\theta^*(u_1, u_2)} \leqq \frac{u_2 - u_1}{\theta^*(u_1, u_2)} \sup_{u_1 \leqq u \leqq u_2} |\theta'(u)| \leqq k \sup_{u_1 \leqq u \leqq u_2} |\theta'(u)| \to 0,$$

as $u_1 \to +\infty$, by hypothesis. This proves the lemma.

COROLLARY. *Let*

$$0 \leqq x_2 - x_1 \leqq c, \qquad c > 0, \text{ a constant,} \tag{15.2}$$

and $u_1 = \min_{|y| \leqq \pi/2} U(x_1 + iy)$, $u_2 = \max_{|y| \leqq \pi/2} U(x_2 + iy)$. *Then, uniformly for all* x_2, *satisfying* (15.2),

$$\lim_{x_1 \to +\infty} \frac{\theta^*(u_1, u_2)}{\theta_*(u_1, u_2)} = 1.$$

Proof. From Theorem III (a) we have

$$\pi \int_{u_1}^{u_2} \frac{du}{\theta(u)} \leqq x_2 - x_1 + 4\pi \leqq c + 4\pi$$

by (15.2). On the other hand, since $u_2 \geqq u_1$,

$$\int_{u_1}^{u_2} \frac{du}{\theta(u)} \geqq \frac{u_2 - u_1}{\theta^*(u_1, u_2)},$$

and hence $0 \leqq u_2 - u_1 \leqq \{(4\pi + c)/\pi\}\theta^*(u_1, u_2)$. Thus the result follows from Lemma 7

16. **Asymptotic expressions for $Z(w)$ and $Z'(w)$.** We prove now the following

THEOREM X. *If S is an L-strip with boundary inclination 0 at $u = +\infty$, we have*

(i) *For $w = u + iv$ in S, uniformly in v,*

$$\lim_{u \to +\infty} \frac{Z(w)}{\displaystyle\int_{u_0}^{u} \frac{dt}{\theta(t)}} = \pi, \qquad u_0 \text{ defined as in §1 } (a).$$

(ii) *Uniformly in any fixed strip $|y| \leqq \beta < \pi/2$ $(z = x + iy)$*[22],

$$\lim_{x \to +\infty} \frac{|W'(z)|}{\theta(u)} = \frac{1}{\pi} \quad (u = U(z)), \qquad \lim_{x \to +\infty} \frac{\log |W'(z)|}{|z|} = 0.$$

(iii) *The straight line Λ_y: $y = \text{const.}$, $|y| < \pi/2$, is mapped by $Z(w)$ onto a curve L_y which for sufficiently large u is represented by an equation of the form*

$$v = f_y(u) \equiv \psi(u) + \frac{\theta(u)}{\pi} y + o[\theta(u)] \qquad \text{as } u \to +\infty, \tag{16.1}$$

uniformly in $|y| < \pi/2$[23].

(iv) *Let the image of the strip*

$$\textstyle\sum_\alpha: \; \{-\infty < x < \infty, \; |y| \leqq \alpha\}, \qquad 0 < \alpha < \pi/2,$$

by means of $w = W(z)$ be T_α, and let S_β be the region

$$\left\{ u > u_0, \left| \frac{v - \psi(u)}{\theta(u)} \right| < \frac{\beta}{\pi} \right\}, \qquad 0 < \beta < \frac{\pi}{2}.$$

If $0 < \alpha \pm \epsilon < \pi/2$, $\epsilon > 0$, then there exists an $N = N(\epsilon; \alpha)$ such that the part of T_α which lies in $u \geqq N$, contains that part of $S_{\alpha-\epsilon}$ which is in $u \geqq N$ and is contained in $S_{\alpha+\epsilon}$.

---

[22] Since $\arg W'(z) \to 0$ as $x \to +\infty$, the first of these relations implies that even

$$\lim_{x \to +\infty} \frac{W'(z)}{\theta(u)} = \frac{1}{\pi}.$$

[23] This implies the following fact: If Λ is a curve in the strip $|y| < \pi/2$ which approaches the line Λ_t as an asymptote, as $x \to +\infty$, then, for any point (u, v) on the image L of Λ in the w-plane, $v = \psi(u) + (\theta(u)/\pi)t + o[\theta(u)]$ as $u \to +\infty$. (L need not be representable in the form $v = f(u)$.) For, if $\epsilon > 0$ is given, there exists an x_1 such that the part of Λ for $x \geqq x_1$ lies "between" the lines $\Lambda_{t-\epsilon}$ and $\Lambda_{t+\epsilon}$ (if $|t \pm \epsilon| < \pi/2$; in case $t = \pi/2$ or $t = -\pi/2$, it lies "between" $\Lambda_{(\pi/2)-\epsilon}$ and $\Lambda_{\pi/2}$ or "between" $\Lambda_{-(\pi/2)}$ and $\Lambda_{-(\pi/2)+\epsilon}$, respectively). The image of that part of Λ lies in a certain half-plane $u \geqq u_1$ "between" the curves $L_{t-\epsilon}$ and $L_{t+\epsilon}$ (or $L_{(\pi/2)-\epsilon}$, $L_{\pi/2}$ or $L_{-\pi/2}$, $L_{-(\pi/2)+\epsilon}$, respectively). Since ϵ may be taken arbitrarily small, this proves the above assertion.

(v) *In any region S_β, $0<\beta<\pi/2$, β fixed,*

$$\lim_{u\to+\infty}\left[\,|Z'(w)|\,\theta(u)\right] = \pi, \qquad w = u + iv,\ w \in S_\beta. \tag{16.2}$$

REMARK. Part (iii) may be considered as an extension in a *certain direction* of Carathéodory's well known result, which states that the map of the interior of a closed Jordan curve Γ onto the circle $|\zeta|<1$ is *"quasi-conformal"* at a boundary point ω_0 with a *corner of measure* $\beta>0$, i.e., that the angles at corresponding boundary points ω_0 and ζ_0 are transformed proportionally([24]). Transformation of the interior of Γ onto a strip S by means of the function $w=\log[1/(\omega-\omega_0)]$ and of $|\zeta|<1$ onto the strip $|y|<\pi/2$, in such a way that $\zeta=\zeta_0$ corresponds to $x=+\infty$, leads to the following statement of Carathéodory's theorem: Let C be a closed Jordan curve through $w=\infty$ which, in a neighborhood of $w=\infty$, consists of two branches C_+ and C_- having the lines $v=\phi_+$ and $v=\phi_-$, respectively, for asymptotes as $u\to+\infty$ $(\phi_+-\phi_-=\theta>0)$. If $W(z)=U(z)+iV(z)$ maps $|y|<\pi/2$ onto the interior S of C and if $\lim_{x\to+\infty} U(z)=+\infty$, then

$$V(z) = \tfrac{1}{2}[\phi_+ + \phi_-] + \frac{\theta}{\pi}\,y + o(1) \qquad \text{as } x \to +\infty, \tag{16.3}$$

uniformly in $|y|<\pi/2$.

If S is a simple Jordan strip, the hypothesis regarding the asymptotes of C_+ and C_- means that

$$\lim_{u\to+\infty}\phi_+(u) = \phi_+ \quad\text{and}\quad \lim_{u\to+\infty}\phi_-(u) = \phi_- \text{ exist.}$$

Similarly, Part (ii) can be considered as an extension in a certain direction of a theorem of Ostrowski which (formulated for infinite strips) states: Under the hypothesis of Carathéodory's theorem

$$\lim_{x\to+\infty}|W'(z)| = \frac{\theta}{\pi}, \qquad \lim_{x\to+\infty}\left[\frac{\log|W'(z)|}{|z|}\right] = 0,$$

uniformly in any strip $|y|\leqq\beta<\pi/2$. Here θ may be greater than or equal to 0([25]).

([24]) Carathéodory [1, pp. 40–41] and [2, pp. 19–93]. See also Lindelöf [1, p. 87]. Carathéodory's theorem has been generalized in two other directions: 1. The assumption that C_+ and C_- approach distinct asymptotes has been replaced by the condition that C_+ and C_- "oscillate" within the strips $\phi_+-k_+\leqq v\leqq\phi_++k_+$ and $\phi_--k_-\leqq v\leqq\phi_-+k_-$ respectively (k_+, k_- constants). 2. It has been extended to the case where C_+ and C_- may be arbitrary continua forming the boundary of S (and not necessarily Jordan curves). For these extensions see Gross, [1, p. 278], Ostrowski [1, pp. 172–174], Wolff [2, p. 42] and [3, p. 46], Warschawski [2, p. 674]. Ostrowski [2, p. 77] gives a necessary and sufficient condition in order that the mapping function of a simply-connected region onto a half-plane preserve angles at an accessible boundary point.

([25]) Ostrowski [1]; the first of these equations is his relation (11.3) of page 101, the second follows, for $\theta>0$, by combination of his relations (62.2) and (62.1) of page 174, and, for $\theta=0$, by combination of his relations (68.7) of page 185 and (62.1) of page 174.

Both of these theorems do not require that C_+ and C_- have an L-tangent at $u=+\infty$, as our Theorem X does. Our extension, however, refers to the fact that (iii) might be substituted for (16.3) when C_+ and C_- have *no* asymptotes and that in (ii) the *asymptotic behavior* of $|W'(z)|$ at $x=+\infty$ is given, whether $\theta(u)=\phi_+(u)-\phi_-(u)$ approaches a limit as $u\to+\infty$ or not.

17. **Proof of Theorem** X. (i) By hypothesis, $\psi'(u)$ and $\theta'(u)$ approach 0 as $u\to+\infty$. Hence it follows from Theorem IV (b) that there exists, for every $\epsilon>0$, an $N_1(\epsilon)\geqq u_0$ such that for all $w=u+iv$ and $w_1=u_1+iv_1$ in S for which $u\geqq u_1\geqq N_1(\epsilon)$:

$$X(w)-X(w_1)\leqq \pi(1+\epsilon)\int_{u_1}^{u}\frac{dt}{\theta(t)}+\epsilon\leqq \pi(1+\epsilon)\int_{u_0}^{u}\frac{dt}{\theta(t)}+\epsilon.$$

Hence,

$$\frac{X(w)-X(w_1)}{\displaystyle\int_{u_0}^{u}\frac{dt}{\theta(t)}}\leqq \pi(1+\epsilon)+\frac{\epsilon}{\displaystyle\int_{u_0}^{u}\frac{dt}{\theta(t)}}\cdot$$

Keeping here w_1 fixed and letting $u\to+\infty$ we find, since $\int_{u_0}^{u}[\theta(t)]^{-1}dt\to\infty$ as $u\to+\infty$,

$$\limsup_{u\to+\infty}\frac{X(w)}{\displaystyle\int_{u_0}^{u}\frac{dt}{\theta(t)}}\leqq \pi(1+\epsilon),\qquad \text{uniformly with respect to } v.$$

Since the left-hand side of this inequality is *independent* of ϵ, we may let $\epsilon\to 0$ and find

$$\limsup_{u\to+\infty}\frac{X(w)}{\displaystyle\int_{u_0}^{u}\frac{dt}{\theta(t)}}\leqq \pi. \tag{17.1}$$

On the other hand, by Theorem III (a), uniformly with respect to v,

$$\liminf_{u\to+\infty}\frac{X(w)}{\displaystyle\int_{u_0}^{u}\frac{dt}{\theta(t)}}\geqq \pi, \tag{17.2}$$

and (17.1) and (17.2) together imply that, uniformly in v,

$$\lim_{u\to\infty}\frac{X(w)}{\displaystyle\int_{u_0}^{u}\frac{dt}{\theta(t)}}=\pi. \tag{17.3}$$

The result (i) follows now from (17.3), if we observe that $Z(w)=X(w)+iY(w)$ and that $|Y(w)|\leqq\pi/2$, while $\int_{u_0}^{u}[\theta(t)]^{-1}dt\to\infty$ as $u\to+\infty$.

(ii) Let L denote the image in the w-plane of the real axis of the z-plane by means of $w=W(z)$. Let $z_n=3n$ $(n=1, 2, 3, \cdots)$ and $w_n=W(z_n)=u_n+iv_n$. Observing that $\lim_{u\to+\infty}\psi'(u)=\lim_{u\to+\infty}\theta'(u)=0$, we have, by Theorem IV (b), for all sufficiently large n,

$$3 = 3(n+1) - 3n \leqq 2\pi\int_{u_n}^{u_{n+1}}\frac{du}{\theta(u)} + 1,$$

or

$$\pi\int_{u_n}^{u_{n+1}}\frac{du}{\theta(u)} \geqq 1. \tag{17.4}$$

Let $\epsilon>0$ be given. By Theorem IV (b) there exists an $n_0(\epsilon)$ such that for $n\geqq n_0(\epsilon)$

$$\begin{aligned} X(w_{n+1}) - X(w_n) &\leqq \pi\left(1+\frac{\epsilon}{2}\right)\int_{u_n}^{u_{n+1}}\frac{du}{\theta(u)} + \frac{\epsilon}{2} \\ &\leqq \pi(1+\epsilon)\int_{u_n}^{u_{n+1}}\frac{du}{\theta(u)}, \end{aligned} \tag{17.5}$$

because of (17.4). Similarly, by Theorem III (b), there is an $n_1(\epsilon)\geqq n_0(\epsilon)$ such that for $n\geqq n_1(\epsilon)$:

$$X(w_{n+1}) - X(w_n) \geqq \pi\int_{u_n}^{u_{n+1}}\frac{du}{\theta(u)} - \frac{\epsilon}{1+\epsilon} \geqq \frac{\pi}{1+\epsilon}\int_{u_n}^{u_{n+1}}\frac{du}{\theta(u)}. \tag{17.6}$$

Since arg $W'(z)\to 0$ as $x\to+\infty$, the curve L can be represented for sufficiently large u in the form $v=f(u)$ where $f'(u)$ exists and approaches 0 as $u\to+\infty$. Evidently, for all sufficiently large n, $w_n=u_n+if(u_n)$. The functions $X(u+if(u))$ and $\int_{u_n}^{u}[\theta(t)]^{-1}dt$ are differentiable for $u_n\leqq u\leqq u_{n+1}$, and $d/du\int_{u_n}^{u}[\theta(t)]^{-1}dt>0$ for $u\geqq u_n$. Hence, by the extended mean value theorem, for sufficiently large n, say $n\geqq n_2\geqq n_1$:

$$\frac{X(w_{n+1}) - X(w_n)}{\displaystyle\int_{u_n}^{u_{n+1}}\frac{du}{\theta(u)}} = \theta(u_n')\left[\frac{\partial X}{\partial u}\bigg|_{w_n'} + \frac{\partial X}{\partial v}\bigg|_{w_n'} f'(u_n')\right],$$

where $u_n<u_n'<u_{n+1}$ and $w_n'=u_n'+if(u_n')$. Thus, we obtain from (17.5) and (17.6):

$$\frac{\pi}{1+\epsilon} \leqq \theta(u_n')\left[\frac{\partial X}{\partial u}\bigg|_{w_n'} + \frac{\partial X}{\partial v}\bigg|_{w_n'} f'(u_n')\right] \leqq \pi(1+\epsilon).$$

Since

$$\frac{\partial X}{\partial u} = |Z'(w)| \cos(\arg Z'(w)), \qquad \frac{\partial X}{\partial v} = -|Z'(w)| \sin(\arg Z'(w))$$

and

$$\lim_{u\to+\infty} [\arg Z'(w)] = \lim_{u\to+\infty} f'(u) = 0,$$

we have, for all sufficiently large n, say $n \geqq n_3 \geqq n_2$:

$$\frac{\pi}{(1+\epsilon)^2} \leqq \theta(u_n')\,|Z'(w_n')| \leqq \pi(1+\epsilon)^2.$$

Let $z_n' = Z(w_n')$. Evidently $3n \leqq z_n' \leqq 3(n+1)$. Since $Z'(w_n') = 1/W'(z_n')$ we have also

$$\frac{1}{\pi(1+\epsilon)^2} \leqq \frac{|W'(z_n')|}{\theta(u_n')} \leqq \frac{(1+\epsilon)^2}{\pi}.$$

Now let $z = x+iy$ be a point in the strip $|y| \leqq \beta$ with $x \geqq 3n_3$, and let n be such that $3n \leqq x < 3(n+1)$. Since then $|z_n' - x| \leqq 3$, we infer from Theorem II (b) that, uniformly for $|y| \leqq \beta$,

$$\lim_{z\to\infty} \left| \frac{W'(z)}{W'(z_n')} \right| = 1.$$

Moreover, by the corollary of Lemma 7, we have for $u = U(z)$, $3n \leqq x < 3(n+1)$,

$$\lim_{u\to+\infty} \frac{\theta(u)}{\theta(u_n')} = 1.$$

Hence, if x is sufficiently large,

$$\frac{1}{\pi(1+\epsilon)^3} \leqq \frac{|W'(z)|}{\theta(u)} = \left| \frac{W'(z)}{W'(z_n')} \right| \frac{|W'(z_n')|}{\theta(u_n')} \frac{\theta(u_n')}{\theta(u)} \leqq \frac{(1+\epsilon)^3}{\pi},$$

and this proves the first relation of part (ii) of the theorem.

To prove the second relation we use the first one:

$$\log|W'(z)| = \log\theta(u) - \log\pi + \log(1+\delta), \qquad \lim_{z\to+\infty}\delta = 0, \tag{17.7}$$

uniformly in $|y| \leqq \beta$. Now, by part (i), uniformly in $|y| < \pi/2$, as $x \to +\infty$,

$$\frac{\log\theta(u)}{|z|} \sim \frac{\displaystyle\int_{u_1}^{u} \frac{\theta'(t)}{\theta(t)}\,dt + \log\theta(u_1)}{\displaystyle\pi\int_{u_1}^{u} \frac{dt}{\theta(t)}}. \tag{17.8}$$

Here u_1 may be chosen arbitrarily large, but fixed. Let $\epsilon>0$ be given. We choose u_1 such that $|\theta'(t)|<\epsilon$ for $t\geqq u_1$. Then by the mean value theorem

$$\frac{\left|\int_{u_1}^{u}\frac{\theta'(t)}{\theta(t)}dt\right|}{\int_{u_1}^{u}\frac{dt}{\theta(t)}}=|\theta'(\xi)|<\epsilon, \qquad u_1<\xi<u.$$

Keeping u_1 *fixed* and letting $u\rightarrow+\infty$, and hence $x\rightarrow+\infty$, we find from (17.8)

$$\limsup_{x\rightarrow+\infty}\left|\frac{\log\theta(u)}{x}\right|\leqq\epsilon,$$

and the result follows from (17.7).

(iii) a. We shall first prove the *weaker* statement that (16.1) holds uniformly for all y with $|y|\leqq\beta$, β being any fixed positive number less than $\pi/2$.

The image L_y of Λ_y is represented in parametric form by the equations $u=U(x+iy)$, $v=V(x+iy)$, where x is the parameter, $-\infty<x<\infty$, and y is fixed. Evidently

$$V(x+iy)=V(x)+\int_0^y\frac{\partial V(x+i\eta)}{\partial\eta}d\eta. \tag{17.9}$$

For $\zeta=x+i\eta$, $|\eta|\leqq|y|$,

$$\frac{\partial V(x+i\eta)}{\partial\eta}=\Re\left\{\frac{dW(\zeta)}{d\zeta}\right\}=|W'(\zeta)|\cos(\arg W'(\zeta)),$$

and by Theorem II (a), the first of the relations (ii), and the corollary of Lemma 7, we have for $u=U(x)$

$$\frac{\partial V(\zeta)}{\partial\eta}=\frac{\theta(u)}{\pi}(1+o(1)) \qquad \text{as } u\rightarrow+\infty,$$

uniformly for $|\eta|\leqq\beta$. Therefore,

$$V(x+iy)=V(x)+\frac{\theta(u)}{\pi}y+o[\theta(u)], \qquad \text{as } x\rightarrow+\infty, \tag{17.10}$$

uniformly for $|y|\leqq\beta$.

Since $\arg W'(z)\rightarrow 0$ as $x\rightarrow+\infty$, uniformly in $|y|<\pi/2$, there exists an N' such that for $u\geqq N'$ every curve L_y, $|y|<\pi/2$, is representable in the form $v=f_y(u)$, where $df_y(u)/du$ is continuous and approaches 0 as $u\rightarrow+\infty$, uniformly for $|y|<\pi/2$. Let $u=U(x)$, $u'=U(x+iy)$ and let x be so large that u, $u'\geqq N''\geqq N'$, where N'' is so chosen that $|df_y(u)/du|<1$ for $u\geqq N''$. Hence[28]

[28] $\theta^*(u, u')=\max\theta(t)$ for t between u and u'.

$$|f_y(u) - V(x+iy)| = |f_y(u) - f_y(u')| < |u - u'| \leqq \theta^*(u, u') \left| \int_u^{u'} \frac{dt}{\theta(t)} \right|.$$

By Theorem III (b), $\int_u^{u'} [\theta(t)]^{-1} dt = o(1)$ and, by the Corollary of Lemma 7, $\theta^*(u, u')/\theta(u) \to 1$ as $u \to +\infty$. We find therefore

$$|f_y(u) - V(x+iy)| \leqq \theta(u)\, o(1), \qquad \text{as } u \to +\infty,$$

uniformly in $|y| < \pi/2$. Hence we obtain from (17.10)

$$(17.11) \qquad f_y(u) = V(x) + \frac{\theta(u)}{\pi} y + o[\theta(u)] \qquad \text{as } u \to +\infty,$$

uniformly for $|y| \leqq \beta$.

Let now an ϵ, $0 < \epsilon < \pi/8$, be assigned. We may apply (17.11) for the particular value of $\beta = \pi/2 - \epsilon$. Let the curves L_y for $y = \pi/2 - \epsilon$ and $y = -\pi/2 + \epsilon$, for sufficiently large u, be represented by the equations $v = f_+(u)$ and $v = f_-(u)$, respectively. Then it follows from (17.11) that

$$f_+(u) = V(x) + \theta(u)\left[\frac{1}{2} - \frac{\epsilon}{\pi}\right] + o[\theta(u)],$$

$$f_-(u) = V(x) + \theta(u)\left[\frac{\epsilon}{\pi} - \frac{1}{2}\right] + o[\theta(u)]$$

as $u \to +\infty$. Addition and subtraction of these expressions give

$$(17.12) \qquad \begin{aligned} V(x) &= \frac{f_+(u) + f_-(u)}{2} + o[\theta(u)], \\ f_+(u) - f_-(u) &= \theta(u)\left[1 - \frac{2\epsilon}{\pi}\right] + o[\theta(u)], \end{aligned}$$

respectively. Substituting $\theta(u) = \phi_+(u) - \phi_-(u)$ into the second expression we find

$$[\phi_+(u) - f_+(u)] + [f_-(u) - \phi_-(u)] = \frac{2\epsilon}{\pi}\theta(u) + o[\theta(u)].$$

Since each of the summands on the left-hand side is greater than or equal to 0,

$$0 \leqq \phi_+(u) - f_+(u) \leqq \left[\frac{2\epsilon}{\pi} + o(1)\right]\theta(u),$$

$$0 \leqq f_-(u) - \phi_-(u) \leqq \left[\frac{2\epsilon}{\pi} + o(1)\right]\theta(u).$$

Hence we obtain from the first of the relations (17.12)

$$(17.13)\qquad V(x) = \frac{\phi_+(u) + \phi_-(u)}{2} + \left[\lambda \frac{2\epsilon}{\pi} + o(1)\right]\theta(u)$$

for a suitable λ, $|\lambda| \leqq 1$. Substitution of (17.13) into (17.11) gives

$$f_y(u) = \tfrac{1}{2}(\phi_+(u) + \phi_-(u)) + \frac{\theta(u)}{\pi}\cdot y + \left[\frac{2\epsilon}{\pi}\lambda + o(1)\right]\theta(u) \quad \text{as } u \to +\infty,$$

uniformly for $|y| \leqq \beta$. This proves our result in the above stated weaker form, since it shows that, for all $|y| \leqq \beta$,

$$(17.14)\qquad \left| f_y(u) - \psi(u) - \frac{\theta(u)}{\pi}\, y \right| \leqq \left[\frac{2\epsilon}{\pi} + o(1)\right]\theta(u) < \epsilon\theta(u),$$

for all sufficiently large u, say $u \geqq N = N(\epsilon; \beta) \geqq N''$.

b. Using this (weaker) result *only* we shall prove part (iv) of our theorem below. Anticipating here part (iv), we can readily prove (16.1) in the complete form, namely that $o[\theta(u)]$ in (16.1) holds uniformly for $|y| < \pi/2$.

Let ϵ, $0 < \epsilon < \pi/8$, be given. Choose $\beta = \pi/2 - \epsilon/2$ and determine the index $N(\epsilon; \pi/2 - \epsilon/2)$ so that (17.14) holds. Let $\pi/2 - \epsilon/2 < y < \pi/2$. Then, by part (iv), there exists an $N_1(\epsilon) \geqq N$ such that all points $w = u + if_y(u)$ with $u \geqq N_1$ lie in the exterior of $S_{\beta - \epsilon/2}$, i.e., either

$$\psi(u) + \frac{\theta(u)}{\pi}\left(\frac{\pi}{2} - \epsilon\right) < f_y(u) < \phi_+(u)$$

or

$$\phi_-(u) < f_y(u) < \psi(u) + \frac{\theta(u)}{\pi}\left(\epsilon - \frac{\pi}{2}\right).$$

Since we assumed $\pi/2 - \epsilon/2 < y < \pi/2$ and $\lim_{y \to \pi/2} f_y(u) = \phi_+(u)$, the first inequality holds. Similarly, if $-\pi/2 < y < -\pi/2 + \epsilon/2$, the second inequality holds. Hence, for all $u \geqq N_1(\epsilon)$,

$$\left| f_y(u) - \psi(u) - \frac{\theta(u)}{\pi}\, y \right| < \epsilon\theta(u) \qquad \text{for all } |y| < \pi/2.$$

(iv) Let ϵ be given as stated in the theorem. By (17.14) there exists an $N = N(\epsilon/\pi; \alpha)$ such that all points $w \in T_\alpha$ with $\Re w \geqq N$ are contained in the region

$$\psi(u) - \frac{\theta(u)}{\pi}\alpha - \frac{\theta(u)}{\pi}\epsilon < v < \psi(u) + \frac{\theta(u)}{\pi}\alpha + \frac{\theta(u)}{\pi}\epsilon, \qquad u \geqq N.$$

This proves that the part of T_α which lies in $u \geqq N$ is contained in $S_{\alpha+\epsilon}$. In a similar manner it may be seen that it contains the part of $S_{\alpha-\epsilon}$ within $u \geqq N$.

(v) Relation (16.2) follows immediately from the first of the relations (ii) and part (iv) if we observe that, for $w = W(z)$, $Z'(w) = 1/W'(z)$.

18. Corollaries of Theorem X. We now prove the following corollaries.

COROLLARY 1. *If S is an L-strip with boundary inclination γ, $|\gamma|<\pi/2$ at $u=+\infty$, then parts* (iii) *and* (iv) *of Theorem* X *remain unchanged and parts* (ii) *and* (v) *are to be replaced by the relations*

(ii*) $\lim_{x\to+\infty} |W'(z)|/\theta(u)=(\cos\gamma)/\pi$, $\lim_{x\to+\infty} [\log |W'(z)|]/|z|=0$, *uniformly in y,*

(v*) $\lim_{u\to+\infty} [|Z'(w)|\theta(u)]=\pi/\cos\gamma$, *uniformly for $w=u+iv\in S_\beta$, $0<\beta<\pi/2$, respectively. Moreover, uniformly for $w\in S$,*

$$Y(w) = \pi\frac{v-\psi(u)}{\theta(u)} + o(1) \qquad \textit{as } u\to+\infty. \tag{18.1}$$

To prove the corollary we rotate the coordinate system through the angle γ in the positive direction, thus obtaining a new set of coordinates $(\bar u, \bar v)$, $\bar u+i\bar v=\bar w$, where $\bar w=e^{-i\gamma}w$. Let, for sufficiently large $\bar u$, $\bar v=\bar\phi_+(\bar u)$ and $\bar v=\bar\phi_-(\bar u)$ represent C_+ and C_-, respectively. Consider a point C: $w=u+iv\in S$ and let $\bar w=e^{-i\gamma}w$. Denote the points $u+i\phi_-(u)$, $u+i\phi_+(u)$, in the w-plane, and $\bar u+i\bar\phi_-(\bar u)$, $\bar u+i\bar\phi_+(\bar u)$ in the $\bar w$-plane by A, B, $\bar A$, $\bar B$ respectively. In the triangle $\bar AAC$, the angles $\bar A$ and A approach $\pi/2$ and $\pi/2-|\gamma|$ respectively, and in $\bar BBC$, the angles $\bar B$ and B approach $\pi/2$ and $\pi/2-|\gamma|$ respectively as $u\to+\infty$, *uniformly* with respect to the position of C on AB. Hence

$$\frac{AC}{\bar AC}=\frac{\sin\bar A}{\sin A}\to\frac{1}{\cos\gamma}, \qquad \frac{BC}{\bar BC}\to\frac{1}{\cos\gamma} \qquad \text{as } u\to+\infty. \tag{18.2}$$

and therefore $(\bar\theta(\bar u)=\bar\phi_+(\bar u)-\bar\phi_-(\bar u))$

$$\theta(u) = AB = AC+BC = \frac{\overline{AB}}{\cos\gamma} + o(\overline{AB}) = \frac{\bar\theta(\bar u)}{\cos\gamma}(1+o(1)) \qquad \text{as } u\to+\infty, \tag{18.3}$$

the convergence in (18.2) and (18.3) being *uniform* with respect to the position of C on AB.

Now (ii*) follows immediately if we observe that for $\bar W(z)=e^{-i\gamma}W(z)$ by Theorem X (ii) and subsequently by (18.3), uniformly for $|y|\leqq\beta$:

$$\frac{1}{\pi} = \lim_{x\to+\infty}\frac{|\bar W'(z)|}{\bar\theta(\bar u)} = \lim_{x\to+\infty}\frac{|W'(z)|}{\theta(u)\cos\gamma}, \qquad \bar u=\Re\bar W(z);\ u=\Re W(z),$$

and that

$$\frac{\log|\bar W'(z)|}{|z|}=\frac{\log|W'(z)|}{|z|}.$$

To prove now that part (iii) of Theorem X remains unchanged, we note that in the figure which was used above, $AC=v-\phi_-(u)$, $BC=\phi_+(u)-v$,

$\overline{AC}=\bar{v}-\bar{\phi}_-(\bar{u})$, $\overline{BC}=\bar{\phi}_+(\bar{u})-\bar{v}$. From (18.2) we obtain therefore (leaving off the arguments u, $\bar{u}$)

$$(18.4)\quad v-\phi_- = (\bar{v}-\bar{\phi}_-)\frac{1}{\cos\gamma}+o(\bar{\theta}),\quad \phi_+-v=(\bar{\phi}_+-\bar{v})\frac{1}{\cos\gamma}+o(\bar{\theta}),$$

as $u\to+\infty$, *uniformly* with respect to the position of C on AB. By Theorem X (iii), there exists an $\overline{N}$ (independent of y), such that any line Λ_y: $y=$const. ($|y|<\pi/2$) is mapped by $\bar{w}=\overline{W}(z)$ onto a curve L_y which, for all $\bar{u}\geqq\overline{N}$, can be represented in the form ($\bar{\psi}=\frac{1}{2}[\bar{\phi}_++\bar{\phi}_-]$):

$$\bar{v}=\bar{\psi}+\frac{\bar{\theta}}{\pi}y+o(\bar{\theta})\qquad \text{as } \bar{u}\to+\infty, \text{ uniformly for } |y|<\pi/2.$$

Hence, for $\bar{w}=\bar{u}+\bar{v}i$ on L_y ($\bar{u}\geqq\overline{N}$), uniformly for $|y|<\pi/2$,

$$\bar{v}-\bar{\phi}_-=\frac{\bar{\theta}}{\pi}\left[\frac{\pi}{2}+y+o(1)\right],\qquad \bar{\phi}_+-\bar{v}=\frac{\bar{\theta}}{\pi}\left[\frac{\pi}{2}-y+o(1)\right]$$

as $u\to+\infty$. Substituting these values for $\bar{v}-\bar{\phi}_-$ and $\bar{\phi}_+-\bar{v}$ into the first and second equations of (18.4), respectively and using (18.3) we find that there exists an N (independent of y) such that any point $w=u+iv$ on L_y with $u\geqq N$ satisfies the relations

$$(18.5)\quad v-\phi_-=\frac{\theta}{\pi}\left[\frac{\pi}{2}+y+o(1)\right],\qquad \phi_+-v=\frac{\theta}{\pi}\left[\frac{\pi}{2}-y+o(1)\right],$$

as $u\to+\infty$, uniformly in $|y|<\pi/2$. Subtraction of the second equation in (18.5) from the first gives

$$(18.6)\qquad v=\psi+\frac{\theta}{\pi}y+o(\theta)\qquad \text{as } u\to+\infty, \text{ uniformly in } |y|<\pi/2.$$

It may be shown now by use of (18.6) (as in the proof of part (iv) of Theorem X) that the statement of part (iv) remains unchanged in the case of any γ, $|\gamma|<\pi/2$. Part (v*) of the corollary follows then from part (ii*) in the same manner as part (v) of the theorem follows from part (ii).

Finally, (18.1) is obtained by solving (18.6) for y, since for $w=u+iv$ on L_t: $Y(w)=t$.

COROLLARY 2. *Let S be an L-strip with the boundary inclination γ, $|\gamma|<\pi/2$, at $u=+\infty$.*

(i) *If $\gamma=0$, then for $z_1=x_1+iy_1$, $z_2=x_2+iy_2$, $|x_2-x_1|\leqq M$ ($M=$const.),*

$$(18.7)\qquad U(z_2)-U(z_1)=\frac{\theta(u)}{\pi}(x_2-x_1)+o[\theta(u)]$$

uniformly for $|y| < \pi/2$ *as* $x_1, x_2 \to +\infty$. *Here* u *may be taken equal to any number between* u_1 *and* u_2.

In particular, if $\lim_{u\to+\infty} \theta(u) = \theta$ *exists, then* (18.7) *becomes*

$$U(z_2) - U(z_1) = \frac{\theta}{\pi}(x_2 - x_1) + o(1). \tag{18.7*}$$

(ii) *If* $\limsup_{u\to+\infty} \theta(u) = \theta^*$, $\liminf_{u\to+\infty} \theta(u) = \theta_*$, *then, uniformly in* $|y| < \pi/2$,

$$\frac{\theta_* \cos^2 \gamma}{\pi} \leqq \liminf_{x\to+\infty} \frac{U(z)}{x} \leqq \limsup_{x\to+\infty} \frac{U(z)}{x} \leqq \frac{\theta^* \cos^2 \gamma}{\pi}. \tag{18.8}$$

Thus, if $\lim_{u\to+\infty} \theta(u) = \theta$ *exists:*

$$\lim_{x\to+\infty} \frac{U(z)}{x} = \frac{\theta \cos^2 \gamma}{\pi}, \qquad \lim_{x\to+\infty} \frac{W(z)}{z} = \frac{\theta \cos \gamma}{\pi} e^{i\gamma}. \tag{18.8*}$$

REMARK. In the case that

$$\lim_{u\to+\infty} \phi_+(u) = \phi_+, \qquad \lim_{u\to+\infty} \phi_-(u) = \phi_- \text{ exist,} \tag{18.9}$$

both of these parts are known theorems([27]).

Proof. (i) Let $u_1 = U(z_1)$, $u_2 = U(z_2)$, $x_2 \geqq x_1$. By Theorems III (b) and IV (b)

$$-\epsilon_1 + \pi \int_{u_1}^{u_2} \frac{du}{\theta(u)} \leqq x_2 - x_1 \leqq \pi(1 + \epsilon_2) \int_{u_1}^{u_2} \frac{du}{\theta(u)} + \epsilon_3, \tag{18.10}$$
$$\lim_{x_1\to+\infty} \epsilon_k = 0, \quad k = 1, 2, 3.$$

By the corollary of Lemma 7, for any u between u_1 and u_2:

$$\int_{u_1}^{u_2} \frac{dt}{\theta(t)} = \frac{1}{\theta(u)}(u_2 - u_1)(1 + \epsilon_4), \qquad \lim_{x_1\to+\infty} \epsilon_4 = 0.$$

Substituting this value of $\int_{u_1}^{u_2}[\theta(t)]^{-1}dt$ into (18.10) and observing that $|x_2 - x_1| \leqq M$, we find (18.7).

([27]) In the case that $\phi_+ - \phi_- = \theta > 0$, (18.7*) has been proved under the less restrictive hypothesis that S is a region as described in the statement of Carathéodory's theorem in the remark of §16, whose boundary curves C_+ and C_- (which approach the aysmptotes $v = \phi_+$ and $v = \phi_-$ respectively) satisfy an *additional condition* regarding their "*oscillation*" ("Reguläre Unbewalltheit"). See Wolff [1, p. 217], Warschawski [1, p. 326], Ostrowski [1, p. 117, relation (21.2)]. For $\theta = 0$, Ostrowski [1, p. 177] proved (18.7*) under the assumption that S is an L-strip satisfying (18.9). Compare also Ostrowski's extension of the first case ($\theta > 0$) to regions with general boundaries [2, pp. 88, 95].—Part (ii) of Corollary 2, under the assumption (18.9) is due to Ostrowski [1, p. 174, relation (62.1)].

(ii) Assume first that $\gamma=0$. Then it is sufficient to prove that

$$(18.11)\qquad \frac{\theta_*}{\pi} \leqq \liminf_{u\to+\infty} \frac{u}{X(w)} \leqq \limsup_{u\to+\infty} \frac{u}{X(w)} \leqq \frac{\theta^*}{\pi}, \qquad \text{uniformly for } w \in S.$$

Let $\epsilon>0$ be given. Take u_1 so large that

$$\theta_* - \epsilon \leqq \theta(u) \leqq \theta^* + \epsilon \qquad \text{for } u \geqq u_1.$$

By Theorem X (i), uniformly for $w\in S$, as $u\to+\infty$,

$$\frac{u}{X(w)} \sim \frac{u}{\pi\int_{u_1}^{u}\frac{dt}{\theta(t)}} = \frac{u}{u-u_1}\cdot\frac{\theta(\xi)}{\pi}, \qquad u_1<\xi<u.$$

Hence, keeping u_1 fixed, we have

$$\frac{\theta_*-\epsilon}{\pi} \leqq \liminf_{u\to+\infty} \frac{u}{X(w)} \leqq \limsup \frac{u}{X(w)} \leqq \frac{\theta^*+\epsilon}{\pi}.$$

Since ϵ is arbitrary this proves (18.11).

If $\gamma\neq 0$, let $\bar{w}=e^{-i\gamma}w$, $\overline{W}(z)=\overline{U}(z)+i\overline{V}(z)=e^{-i\gamma}W(z)$. Moreover, let $\bar{\theta}(\bar{u})$ be defined in the $\bar{w}$-plane as $\theta(u)$ is in the w-plane. Then, by the part just proved,

$$(18.12)\qquad \frac{1}{\pi}\liminf_{\bar{u}\to+\infty}\bar{\theta}(\bar{u}) \leqq \liminf_{x\to+\infty}\frac{\overline{U}(z)}{x} \leqq \limsup_{x\to+\infty}\frac{\overline{U}(z)}{x} \leqq \frac{1}{\pi}\limsup_{\bar{u}\to+\infty}\bar{\theta}(\bar{u}).$$

Now

$$U(z) = \overline{U}(z)\cos\gamma - \overline{V}(z)\sin\gamma = \overline{U}(z)\cos\gamma\left\{1-\frac{\overline{V}(z)}{\overline{U}(z)}\tan\gamma\right\}.$$

Here $\overline{V}(z)/\overline{U}(z)\to 0$ as $x\to+\infty$, uniformly for $|y|<\pi/2$. Using this relation in connection with (18.3) we obtain (18.8) from (18.12).

VI. Applications

19. **The general result.** We shall now apply our results to the study of the mapping function in a neighborhood of a finite boundary point ω_0 of a region bounded by a closed Jordan curve. First we derive from our results a theorem which deals with a certain general boundary configuration and then apply it to various special cases.

(a) Let R be the interior of a closed Jordan curve Γ in the ω-plane and let $\omega=0$ be on Γ. Suppose that in a neighborhood of $\omega=0$, say $|\omega|\leqq a$, Γ consists of two arcs Γ_+ and Γ_- which are represented in polar coordinates in the form

$$\phi=\Phi_+(\rho),\qquad \phi=\Phi_-(\rho),$$
$$0<\rho\leqq a,\qquad \Phi_+(\rho)<\Phi_-(\rho),$$

respectively, the functions $\Phi_+(\rho)$, $\Phi_-(\rho)$ being continuous in the interval $0<\rho\leqq a$. The region

$$0 < \rho < a, \qquad \Phi_+(\rho) < \phi < \Phi_-(\rho)$$

is contained in R. We set $\Theta(\rho)\equiv\Phi_-(\rho)-\Phi_+(\rho)$ and $\Psi(\rho)=\frac{1}{2}[\Phi_-(\rho)+\Phi_+(\rho)]$.

Suppose that $\Phi_+(\rho)$ and $\Phi_-(\rho)$ are *absolutely continuous in any closed interval within* $0<\rho\leqq a$ *and that* $\rho[d\Phi_+(\rho)/d\rho]$ *and* $\rho[d\Phi_-(\rho)/d\rho]$, *which exist for* $0<\rho\leqq a$ *except possibly for a set of measure* 0, *approach the same limit*, $\tan\gamma$, $|\gamma|<\pi/2$, *as* $\rho\to 0$.

Finally, let $\zeta=\zeta(\omega)$ map R conformally onto the circle $|\zeta-1|<1$ in such a manner that $\omega=0$ corresponds to $\zeta=0$ and let $\omega=\omega(\zeta)$ denote its inverse function.

(b) Logarithmic transformation of R by means of the function[28] $w=\log(1/\omega)$ and of the circle $|\zeta-1|<1$ by means of $z=\log[(2-\zeta)/\zeta]$ gives at once the following results:

THEOREM XI(A). *Under the above stated hypotheses we have*[29]:

(i) *If* $\gamma=0$, *then for any branch of* $\log\zeta(\omega)$, $|\omega|=\rho$, *uniformly in* R,

$$\lim_{\rho\to 0}\frac{\log\zeta(\omega)}{\displaystyle\int_\rho^a\frac{dr}{r\Theta(r)}}=-\pi.$$

(ii) *Uniformly in any angle* $|\arg\zeta|\leqq\beta<\pi/2$, *as* $\zeta\to 0$,

$$\frac{|\omega'(\zeta)|}{\left|\dfrac{\omega(\zeta)}{\zeta}\right|}\sim\frac{\Theta(\rho)}{\pi}\cos\gamma, \qquad |\omega|=\rho.$$

(iii) *Any circular arc* λ_t, *t fixed*, $|t|<\pi/2$: $\{\arg\{(2-\zeta)/\zeta\}=t, |\zeta-1|<1\}$ *is mapped by* $\omega(\zeta)$ *onto a curve* l_t *which in a neighborhood of* $\omega=0$ *is representable in the form*

$$\phi=\Psi(\rho)-\frac{\Theta(\rho)}{\pi}t+o[\Theta(\rho)] \qquad \text{as } \rho\to 0,$$

uniformly for $|t|<\pi/2$.

---

(28) R is transformed by $w=\log 1/\omega$ into an L-strip S whose boundary curves C_+ and C_- are given by the equations $v=\phi_+(u)=-\Phi_+(e^{-u})$, $v=\phi_-(u)=-\Phi_-(e^{-u})$, respectively. To see that C_+ (and similarly C_-) has an L-tangent at $u=+\infty$, one only has to note that

$$\frac{\phi_+(u_2)-\phi_+(u_1)}{u_2-u_1}=-\frac{1}{u_2-u_1}\int_{\rho_1}^{\rho_2}\frac{d\Phi_+}{d\rho}d\rho=-(1+\epsilon)\frac{\tan\gamma}{u_2-u_1}\int_{\rho_1}^{\rho_2}\frac{d\rho}{\rho}$$
$$=-(1+\epsilon)\tan\gamma\frac{\log\rho_2-\log\rho_1}{u_2-u_1}=(1+\epsilon)\tan\gamma,$$

where $\epsilon\to 0$ as $\rho_1, \rho_2\to 0$.

(29) Part (A) follows from Theorem X and Corollary 1 of this theorem.

(iv) *Let σ_α denote the region within $|\zeta-1|<1$ which is bounded by the two circular arcs* $\arg\{(2-\zeta)/\zeta\}=\alpha$ *and* $\arg\{(2-\zeta)/\zeta\}=-\alpha$, $0<\alpha<\pi/2$, *let τ_α be its image in the ω-plane by means of $\omega(\zeta)$, and let s_β $(0<\beta<\pi/2)$ denote the region* $\{0<\rho<a,\ |\phi-\Psi(\rho)|\leqq(\beta/\pi)\Theta(\rho)\}$.

If $0<\alpha\pm\epsilon<\pi/2$, $\epsilon>0$, then there exists an $r=r(\epsilon;\alpha)$ such that the part of τ_α which lies in $0<\rho\leqq r$ contains that part of $s_{\alpha-\epsilon}$ which is in $\rho\leqq r$ and is contained in $s_{\alpha+\epsilon}$.

(v) *Uniformly in any region s_β, $0<\beta<\pi/2$,*

$$\frac{|\zeta'(\omega)|}{\left|\dfrac{\zeta(\omega)}{\omega}\right|}\sim\frac{\pi}{\Theta(\rho)\cos\gamma}\qquad as\ |\omega|=\rho\to 0.$$

(vi) *Uniformly in R, for $\omega=\rho e^{i\phi}$,*

$$\arg\zeta(\omega)=\pi\frac{\phi-\Psi(\rho)}{\Theta(\rho)}+o(1)\qquad as\ \omega\to 0. \tag{19.1}$$

THEOREM XI (B). *If in addition to the hypotheses stated in* §19 (a), *$\rho[d\Phi_+(\rho)/d\rho]$ and $\rho[d\Phi_-(\rho)/d\rho]$ are continuous for $0\leqq\rho\leqq a$ and the integrals*

$$\int_{\rho=0}^{a}|d(\rho\Phi_+'(\rho))|,\quad\int_{\rho=0}^{a}|d(\rho\Phi_-'(\rho))|,\quad\int_0^a\frac{\Theta'^2(\rho)}{\Theta(\rho)}d\rho\ converge, \tag{19.2}$$

then there exists a constant $c>0$ such that, for $\omega=\rho e^{i\phi}$,

$$|\zeta(\omega)|=c\exp\left\{-\pi\int_\rho^a\frac{1+(r\Psi'(r))^2}{r\Theta(r)}dr+\pi\frac{\phi-\Psi(\rho)}{\Theta(\rho)}\tan\gamma+o(1)\right\} \tag{19.3}$$

as $\omega\to 0$ in any way at all in R[30].

(c) REMARK. *If Γ is a curve as described in* §19 (a) *and if, in addition the integrals*

$$\int_0^a\rho\left(\frac{d\Phi_+}{d\rho}\right)^2\frac{d\rho}{\Theta(\rho)}\quad and\quad\int_0^a\rho\left(\frac{d\Phi_-}{d\rho}\right)^2\frac{d\rho}{\Theta(\rho)}\quad converge, \tag{19.4}$$

then $\gamma=0$[31] *and the integral $\int_0^a(\rho\Psi'^2(\rho)/\Theta(\rho))d\rho$ exists, and therefore by the corollary of Theorem* IV (§7), (19.3) *reduces to*

$$|\zeta(\omega)|=c'\exp\left\{-\pi\int_\rho^a\frac{dr}{r\Theta(r)}+o(1)\right\}\quad as\ \omega\to 0\ in\ R\ (c'=const.). \tag{19.5}$$

(d) COROLLARY. *Let R be a region in the ω-plane satisfying the hypotheses*

(30) Part (B) follows from the corollary to Theorem VIII (§14).

(31) If γ were not 0, the convergence of either of the integrals (19.4) would imply that of $\int_0^a[r\Theta(r)]^{-1}dr$, and this integral diverges.

of §19 (a). *If in addition,* $\lim_{\rho\to 0}\Theta(\rho)=\theta$ *exists, then we have*[32]:

(i) *If* $\gamma=0$, *for* ζ_1 *and* ζ_2 *in* $|\zeta-1|<1$, *for which* $0<c_1<|\zeta_1/\zeta_2|<c_2$ (c_1, c_2 *consts.*),

$$\left|\frac{\omega(\zeta_2)}{\omega(\zeta_1)}\right| = \left|\frac{\zeta_2}{\zeta_1}\right|^{\theta/\pi}(1+o(1)) \qquad \text{as } \zeta_1, \zeta_2 \to 0,$$

uniformly in the circle $|\zeta-1|<1$.

(ii) *As* $\zeta\to 0$ *in* $|\zeta-1|<1$ *in any way at all,*

$$\frac{\log|\omega(\zeta)|}{\log|\zeta|}\to\frac{\theta}{\pi}\cos^2\gamma, \qquad \frac{\log\omega(\zeta)}{\log\zeta}\to\frac{\theta}{\pi}e^{i\gamma}\cos\gamma. \tag{19.6}$$

(iii) *As* $\zeta\to 0$ *in any angle* $|\arg\zeta|\leqq\alpha<\pi/2$[33],

$$\frac{\log|\omega'(\zeta)|}{\log|\zeta|}\to\frac{\theta}{\pi}\cos^2\gamma-1. \tag{19.7}$$

We now apply these results to various special cases.

20. **Boundary "elements" with bounded argument oscillation.** Let Γ be a curve as described in §19 (a) for which $\Phi_+(\rho)$ and $\Phi_-(\rho)$ are *bounded for* $0<\rho\leqq a$. Then, necessarily, $\gamma=0$[34], and the results of §19 hold with γ replaced by 0. We consider the case where

$$\lim_{\rho\to 0}\Phi_+(\rho)=\phi_+ \quad\text{and}\quad \lim_{\rho\to 0}\Phi_-(\rho)=\phi_- \quad\text{exist.} \tag{20.1}$$

(a) *Corners.* If $\theta\equiv\phi_--\phi_+>0$, then Γ has an L-corner of measure θ at $\omega=0$ (see §2). In this case parts (ii), (iii) and (vi) of Theorem XI (A) are known, (ii) has been proved by Ostrowski, even under the weaker hypothesis that Γ has a corner at $\omega=0$ (and not necessarily an L-corner), and (iii) and (vi) express the fact that the map is quasi-conformal at $\omega=0$. Furthermore, the results of the corollary for this case as well as those for $\theta=0$ are due to Ostrowski (see [27]).

Formulas (19.3) and (19.5) give *expressions for the order of magnitude of* $|\zeta(\omega)|$. One may ask here, *when, in particular,* $\zeta(\omega)\sim c\omega^{\pi/\theta}$ *as* $\omega\to 0$ *in* $R+\Gamma$ ($c=$ *constant* $\neq 0$). This question has received much attention in the recent lit-

[32] For parts (i) and (ii) see Corollary 2 of Theorem X (§18, (18.7*) and (18.8*)).

[33] The relation (19.7) follows from Theorem XI (A), (ii) and (19.6) if one observes that

$$\log|\omega'(\zeta)| = \log|\omega(\zeta)| - \log|\zeta| + \log\left[\frac{\Theta(\rho)}{\pi}\cos\gamma\right]+o(1)$$

as $\zeta\to 0$ in $|\arg\zeta|\leqq\alpha<\pi/2$.

[34] If γ were not equal to 0, we would have since $(d\Phi_+/d\rho)\rho\to\tan\gamma$, $(\rho_1<\rho_2)$,

$$|\Phi_+(\rho_2)-\Phi_+(\rho_1)| = \left|\int_{\rho_1}^{\rho_2}\frac{d\Phi_+}{d\rho}d\rho\right| \geqq \frac{|\tan\gamma|}{2}\int_{\rho_1}^{\rho_2}\frac{dr}{r}$$

for all sufficiently small ρ_1 and ρ_2. Keeping ρ_2 fixed and letting $\rho_1\to 0$, we would find that $\Phi_+(\rho)$ is not bounded.

erature, especially for the case $\theta=\pi$ where this means that $\zeta(\omega)$ possesses a *nonvanishing derivative at* $\omega=0$. Since by (19.1) $\lim_{\omega\to 0}\arg(\zeta(\omega)/\omega^{\pi/\theta})$ exists, it is sufficient to consider $\lim_{\omega\to 0}|\zeta(\omega)/\omega^{\pi/\theta}|$. The remark of §19 (c) yields the following result *which does not even presuppose any assumption regarding the existence of* $\lim_{\rho\to 0}\theta(\rho)$ *or that of the limits* (20.1).

THEOREM XII. *If* Γ *is a curve as described in* §19 (a) *and if the integrals* (19.4) *converge, then a necessary and sufficient condition that, for some positive* θ, $\lim_{\omega\to 0}|\zeta(\omega)/\omega^{\pi/\theta}|$ *exist for unrestricted approach and be different from* 0 *is that*

$$\int_0^a \frac{\theta-\Theta(\rho)}{\rho\Theta(\rho)}\,d\rho \quad \textit{converge.} \tag{20.2}$$

For, we have, uniformly in R,

$$\left|\frac{\zeta(\omega)}{\omega^{\pi/\theta}}\right| = c'\exp\left\{-\pi\int_\rho^a \frac{dr}{r\Theta(r)}+\frac{\pi}{\theta}\log\frac{a}{\rho}-\frac{\pi}{\theta}\log a+o(1)\right\}$$
$$= c''\exp\left\{-\pi\int_\rho^a \frac{\theta-\Theta(r)}{r\theta\Theta(r)}\,dr+o(1)\right\}$$

as $|\omega|=\rho\to 0$ (c'' is a constant different from 0).

Combining this result with (19.1) we find:

THEOREM XIII. *If* Γ *is a curve as described in* §19 (a) *and if the integrals* (19.4) *converge, then a necessary and sufficient condition in order that for some* $\theta>0$, $\lim_{\omega\to 0}(\zeta(\omega)/\omega^{\pi/\theta})$ *exist for unrestricted approach and be different from* 0 *is that the conditions* (20.1) *and* (20.2) *be satisfied and that* $\phi_- - \phi_+ = \theta$.

The sufficiency of the conditions stated is clear. That (20.2) is necessary follows from Theorem XII. That (20.1) and the relation $\phi_- - \phi_+ = \theta$ are necessary is immediately seen if we observe that $\arg\{\zeta/[\omega(\zeta)]^{\pi/\theta}\}$ approaches a limit as $\zeta\to 0$ in $|\zeta-1|\leq 1$, and then let $\zeta\to 0$, first along the upper and then along the lower semi-circle of $|\zeta-1|=1$.

For $\theta=\pi$ this theorem gives a criterion for the existence of the derivative of $\zeta(\omega)$ at a boundary point. Several criteria (sufficient conditions) for the existence of the angular derivative (i.e., $\lim_{\zeta\to 0}(\omega(\zeta)/\zeta)$ in any fixed angle $|\arg\zeta|\leq\alpha<\pi/2$) are known which apply to even more general types of regions than those bounded by Jordan curves. (For sufficiently smooth boundaries the existence of the derivatives for unrestricted approach can be inferred from that of the angular derivative.) The sharpest of these criteria to date is due to Ahlfors([38]). Since our theorem refers to a smaller class of regions it obviously does not contain that of Ahlfors([39]). On the other hand, the follow-

---

([38]) Ahlfors [1, p. 36].

([39]) However, by use of Theorem XII (B) and a modification of Ahlfors' proof of his criterion, one may obtain a sharper criterion for the existence of the angular derivative.

ing example shows that it also is not contained in Ahlfors' result: Let Γ be a closed Jordan curve through $\omega=0$ and let, in a neighborhood of $\omega=0$ $(0\leq\rho\leq a)$, Γ consist of the two branches Γ_+ and Γ_- represented by

$$\phi = \Phi_+(\rho) \equiv -\frac{\pi}{2} + \delta_1(\rho),\ \phi = \Phi_-(\rho) \equiv \frac{\pi}{2} + \delta_2(\rho),\ \delta_i(\rho) > 0,\ \lim_{\rho\to 0}\delta_i(\rho) = 0,$$

where $\int_0^a[\delta_i(\rho)/\rho]d\rho$ is divergent $(i=1, 2)$, $\int_0^a\{|\delta_1(\rho)-\delta_2(\rho)|/\rho\}d\rho$ convergent, $\rho\delta_i'(\rho)$ continuous, $\rho(\delta_i'(\rho))^2$ integrable, $0\leq\rho\leq a$. It is easily seen that the hypotheses of Theorem XIII are satisfied and that therefore $\lim_{\omega\to 0}[\zeta(\omega)/\omega]$ exists and does not equal 0. However, one of the conditions of Ahlfors' criterion is not satisfied. Suppose Γ is mapped by the function $w=\log(1/\omega)$ onto a strip. The images C_+ and C_- of Γ_+ and Γ_- are represented by the equations

$$v = \phi_+(u) \equiv \frac{\pi}{2} - \delta_1(e^{-u}), \qquad v = \phi_-(u) \equiv -\frac{\pi}{2} - \delta_2(e^{-u}),$$

respectively. Then (in Ahlfors' notation [1, p. 36]) the series $\sum_{\nu=1}^\infty m_\nu$ does not converge. For, if $\nu k\leq u\leq(\nu+1)k$ are the intervals for which the m_ν are formed, then

$$k\sum_{\nu=1}^n m_\nu \geq \sum_{k=1}^n \int_{\nu k}^{(\nu+1)k}\delta_1(e^{-u})du = \int_k^{(n+1)k}\delta_1(e^{-u})du \to +\infty$$

as $n\to\infty$. However, the convergence of this series is one of the conditions of his criterion.

A corollary of Theorem XIII is the following

THEOREM XIV. *Let Γ be a closed Jordan curve through $\omega=0$ which has an L-corner of measure θ at $\omega=0$, formed by the two branches Γ_+ and Γ_-. Suppose that the angles of inclination of the tangents to Γ_+ and Γ_-, considered as functions of ρ, are of bounded variation in a neighborhood of $\omega=0$, $0\leq\rho\leq a$. Let $\zeta(\omega)$ be defined as in* §19 (a). *Then, a necessary and sufficient condition that* $\lim_{\omega\to 0}(\zeta(\omega)/\omega^{\pi/\theta})$ *exist and be different from zero is that the integral* (20.2) *converge.* ($\Phi_+(\rho)$, $\Phi_-(\rho)$ *and* $\Theta(\rho)$ *in* (20.2) *are defined as in* §19 (a).)

Proof. We represent Γ_+ and Γ_- in the form $\phi=\Phi_+(\rho)$ and $\phi=\Phi_-(\rho)$ respectively, $0\leq\rho\leq a$, and show first that the integrals (19.4) converge. Let $\tau_+(\rho)$ denote the angle of inclination of the tangent to Γ_+ at $\omega=\rho e^{i\Phi_+(\rho)}$ (if the tangent exists), $\tau_+(\rho)$ being so chosen that $\lim_{\rho\to 0}\tau_+(\rho)=\phi_+$ (see (20.1)). As is well known,

$$\rho\frac{d\Phi_+(\rho)}{d\rho} = \tan[\tau_+(\rho) - \Phi_+(\rho)].$$

To prove the convergence of the first integral in (19.4) it is sufficient to show that

$$\int_0^a [\tau_+(\rho) - \Phi_+(\rho)] \frac{d\Phi_+(\rho)}{d\rho} d\rho \quad \text{converges.} \tag{20.3}$$

Integration by parts gives ($0 \leqq \epsilon < a$)

$$\begin{aligned} \int_\epsilon^a [\tau_+ - \Phi_+] \frac{d\Phi_+}{d\rho} d\rho &= - \int_{\rho=\epsilon}^a (\Phi_+(\rho) - \phi_+) d[\tau_+(\rho) - \Phi_+(\rho)] \\ &\quad + (\Phi_+(\rho) - \phi_+)(\tau_+(\rho) - \Phi_+(\rho)) \Big|_\epsilon^a \\ &= - \int_{\rho=\epsilon}^a (\Phi_+(\rho) - \phi_+) d\tau_+(\rho) \\ &\quad + \int_{\rho=\epsilon}^a (\Phi_+(\rho) - \phi_+) d[\Phi_+(\rho) - \phi_+] + M(\epsilon, a), \end{aligned}$$

where $M(\epsilon, a)$ is continuous at $\epsilon=0$. Since $\tau_+(\rho)$ is of bounded variation and $\Phi_+(\rho)$ is continuous for $0 \leqq \rho \leqq a$, the first of the last two integrals converges as $\epsilon \to 0$. The second of these integrals has the value $[\frac{1}{2}(\Phi_+(\rho) - \phi_+)^2]_\epsilon^a$ and it approaches, therefore, a finite limit as $\epsilon \to 0$. This proves (20.3).

Similarly it is shown that the second integral in (19.4) converges and Theorem XIV follows now from Theorem XIII.

(b) *Cusps*. Suppose now that Γ has an L-cusp at $\omega=0$. Our main results here are the formulas (19.3), with $\gamma=0$, and (19.5) which give *asymptotic expressions for* $|\zeta(\omega)|$. Parts (iii) and (vi) of Theorem XI (A) take the place of the quasi-conformality in the case of a corner. Part (ii) of this theorem (with $\gamma=0$) is an extension of Ostrowski's result which states that under the weaker assumption that Γ has a cusp at $\omega=0$ (and not necessarily an L-cusp), $\lim_{\zeta \to 0} [\omega'(\zeta)/(\omega(\zeta)/\zeta)] = 0$, in any fixed angle $|\arg \zeta| \leqq \beta < \pi/2$ (cf. the remark to Theorem X, §16). Similarly, part (i) of Theorem XI (A) can be considered as a sharper form of (19.6), since it gives the order of magnitude of $\log |\zeta(\omega)|$ while (19.6) merely states that $(\log |\zeta(\omega)| / \log |\omega|) \to +\infty$ as $\omega \to 0$. The results of the corollary are due to Ostrowski (see ([27])).

Example. It might be of some interest to apply our results to a cusp formed by two arcs Γ_+ and Γ_- which have "*finite order of contact*." Γ_+ and Γ_- are represented in Cartesian coordinates (ξ, η) by the equations

$$\eta = [a + \epsilon(\xi)]\xi^n, \qquad \eta = [b + \delta(\xi)]\xi^m, \qquad 0 \leqq \xi \leqq \xi_0. \tag{20.4}$$

Here a, b, n, m are any real numbers, $n \geqq m > 1$; if $n > m$ then $b > 0$, if $n = m$, $b > a$; $\epsilon(\xi)$, $\delta(\xi)$ have continuous first derivatives, $0 \leqq \xi \leqq \xi_0$ and approach 0 with ξ. It is clear that Γ_+ and Γ_- form an L-cusp so that Theorem XI (A) holds. We examine, therefore, the possibility of applying part (B).

We introduce polar coordinates about $\omega=0$. Using ρ as parameter and writing $\phi=\Phi_+(\rho)$ on Γ_+ and $\phi=\Phi_-(\rho)$ on Γ_- we obtain from (20.4), $\rho>0$,

(20.5) $\rho\sin\Phi_+(\rho)=[a+\epsilon(\xi)]\rho^n\cos^n\Phi_+(\rho),\ \rho\sin\Phi_-(\rho)=[b+\delta(\xi)]\rho^m\cos^m\Phi_-(\rho).$

Hence

$$\sin[\Phi_-(\rho)-\Phi_+(\rho)]=\rho^{m-1}\{(b+\delta)\cos^m\Phi_-\cos\Phi_+-(a+\epsilon)\rho^{n-m}\cos^n\Phi_+\cos\Phi_-\}.$$

Thus we find

$$\Theta(\rho)=\Phi_--\Phi_+\sim b\rho^{m-1}\quad\text{if}\quad n>m;\qquad \Theta(\rho)\sim(b-a)\rho^{m-1}\quad\text{if}\quad n=m,$$

as $\rho\to 0$. Furthermore, differentiation of the first relation of (20.5) with respect to ρ gives

$$\cos\Phi_+(\rho)\frac{d\Phi_+}{d\rho}=(a+\epsilon)(n-1)\rho^{n-2}\{1+o(1)\}\qquad \text{as } \rho\to 0,$$

so that

$$\frac{\rho}{\Theta(\rho)}\left(\frac{d\Phi_+}{d\rho}\right)^2=O(\rho^{n-2})\qquad \text{as } \rho\to 0,$$

and similarly

$$\frac{\rho}{\Theta(\rho)}\left(\frac{d\Phi_-}{d\rho}\right)^2=O(\rho^{m-2}).$$

Hence the conditions (19.4) of the remark of §19 (c) are satisfied and therefore (19.5) holds.

21. **Boundary "elements" with unbounded argument oscillation.** Suppose that Γ is a curve as described in §19 (a) and that $\lim_{\rho\to 0}\Phi_+(\rho)$ and $\lim_{\rho\to 0}\Phi_-(\rho)$ are both $+\infty$ (or $-\infty$). In this case Γ_+ and Γ_- are two "concurrent" spirals having $\omega=0$ as an asymptotic point. Our results of §19 described the behavior of $\zeta(\omega)$ and $\zeta'(\omega)$ as ω approaches the asymptotic point.

However, our methods still apply to a case not included in §19, in which $\rho(d\Phi_+/d\rho)$ and $\rho(d\Phi_-/d\rho)$ both approach $+\infty$ (or $-\infty$). This case is contained in the following more general configuration.

Let R be a simply-connected (single-sheeted) region whose boundary consists of the spirals

$$\Gamma_+:\ \rho=\rho_+(\phi),\qquad \Gamma_-:\ \rho=\rho_-(\phi),\qquad \phi_0\leqq\phi<\infty,$$

and of a Jordan arc connecting the end points $(\phi_0,\ \rho_+(\phi_0))$ of Γ_+ and $(\phi_0,\ \rho_-(\phi_0))$ of Γ_-. It is assumed that $\rho_+(\phi)$ and $\rho_-(\phi)$ are positive and absolutely continuous in any interval $\phi_0\leqq\phi\leqq\phi_1<\infty$, and that $(1/\rho_+(\phi))d\rho_+/d\phi$ and $(1/\rho_-(\phi))d\rho_-/d\rho$ which exist for almost all $\phi\geqq\phi_0$, both approach the limit $-\tan\gamma$, $|\gamma|<\pi/2$ as $\phi\to\infty$. The region

(21.1) $$\phi_0<\phi<\infty,\qquad \rho_+(\phi)<\rho<\rho_-(\phi)$$

is contained in R. We set

$$\Theta(\phi) \equiv \log \frac{1}{\rho_+(\phi)} - \log \frac{1}{\rho_-(\phi)}, \qquad \Psi(\rho) \equiv \frac{1}{2}\left[\log \frac{1}{\rho_+(\phi)} + \log \frac{1}{\rho_-(\phi)}\right].$$

There exists a function $\zeta(\omega)$ which maps R conformally onto the circle $|\zeta-1|<1$ in such a manner that $\lim \zeta(\omega)=0$ as $\arg \omega \to +\infty$, ω being in the region (21.1). The inverse function of $\zeta(\omega)$ will again be denoted by $\omega(\zeta)$.

The function $w=u+iv=i \log (1/\omega)=i \log (1/\rho)+\phi$ maps R onto an L-strip S in the w-plane with the boundary inclination γ at $u=+\infty$. We map the circle $|\zeta-1|<1$ onto the strip $|y|<\pi/2$ by means of the function $z=\log \{(2-\zeta)/\zeta\}$ ($z=0$, when $\zeta=1$). In this way we have again reduced our problem to that of a strip. We can carry over all our theorems on L-strips thus obtaining results on $\zeta(\omega)$ and $\omega(\zeta)$ as $\arg \omega \to +\infty$ for ω in (21.1) and $\zeta \to 0$ in $|\zeta-1|<1$.

To apply Theorem IX and to obtain the analogue of Theorem XI (B) we assume here that the integrals

$$\int_{\phi=\phi_0}^{\infty} \left| d\left(\frac{\rho_+'(\phi)}{\rho_+(\phi)}\right)\right|, \qquad \int_{\phi=\phi_0}^{\infty} \left| d\left(\frac{\rho_-'(\phi)}{\rho_-(\phi)}\right)\right|, \qquad \int_{\phi_0}^{\infty} \frac{\Theta'^2(\phi)}{\Theta(\phi)}\, d\phi$$

converge. There is no difficulty in actually writing out the results obtained for this case.

Bibliography

L. Ahlfors

1. *Untersuchungen zur Theorie der konformen Abbildung und der ganzen Funktionen*, Acta Societatis Scientiarum Fennicae, (n. s.), A, vol. 1 (1930).

C. Carathéodory

1. *Elementarer Beweis für den Fundamentalsatz der konformen Abbildung*, Schwarz-Festschrift, 1914, pp. 19–41.
2. *Conformal Representation*, Cambridge Tracts, 1932.

E. T. Copson

1. *Introduction to the Theory of Functions of a Complex Variable*, Oxford University Press, 1935.

W. Gross

1. *Zum Verhalten analytischer Funktionen in der Umgebung singulärer Stellen*, Mathematische Zeitschrift, vol. 2 (1918), pp. 241–294.

E. Lindelöf

1. *Sur la représentation conforme d'une aire simplement connexe sur l'aire d'un circle*, Compte Rendu du Quatrième Congrès des Mathématiciens Scandinaves, Stockholm, (1916), pp. 59–90.

R. Nevanlinna

1. *Eindeutige analytische Funktionen*, Berlin, 1936.

A. Ostrowski

1. *Über den Habitus der konformen Abbildung am Rande des Abbildungsbereiches*, Acta Mathematica, vol. 64 (1934), pp. 81–185.
2. *Zur Randverzerrung bei konformer Abbildung*, Prace Matematiczno Fiziczne, vol. 44 (1937), pp. 371–471.

T. Radó

1. *Sur la représentation conforme de domaines variables*, Acta Litterarum ac Scientiarum Universitatis Francisco Josephinae, vol. 1 (1923), pp. 180–186.

S. E. Warschawski

1. *Zur Randverzerrung bei konformer Abbildung*, Compositio Mathematica, vol. 1 (1935), pp. 314–343.
2. *On the preservation of angles at a boundary point in conformal mapping*, Bulletin of the American Mathematical Society, vol. 42 (1936), pp. 674–680.

J. Wolff

1. *Sur la représentation conforme des bandes*, Compositio Mathematica, vol. 1 (1934), pp. 217–222.
2. *Sur la conservation des angles dans la représentation conforme d'un domaine au voisinage d'un point frontière*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 200 (1935), pp. 42–43.
3. *Démonstration d'un théorème sur la conservation des angles dans la représentation conforme d'un domaine au voisinage d'un point frontière*, Proceedings, Akademie van Wetenschappen, Amsterdam, vol. 38 (1935), pp. 46–50.

Washington University,
St. Louis, Mo.

ON THE SEMI-CONTINUITY OF DOUBLE INTEGRALS IN PARAMETRIC FORM

BY

TIBOR RADÓ

INTRODUCTION

The purpose of the present investigation is to extend the scope of the important and comprehensive results of McShane on the semi-continuity of double integrals in parametric form(1). The class of surfaces considered by McShane is very general, but we shall see that the conditions placed by him upon his surfaces can be greatly relaxed without hurting the validity of any one of his theorems. Our condition (see 1.19) is however not only less restrictive but also possesses a certain degree of finality. Indeed, we shall be able to prove that *the class of surfaces for which our condition holds is precisely the class of surfaces which admit of a representation where the Lebesgue area of the surface is given by the usual integral formula* (see 3.15, 3.19).

The methods used in this paper are based on previous work by McShane in the calculus of variations and by the author on the area of surfaces. Since this work is scattered in a number of papers in various periodicals, it seemed advisable to attempt at this time a somewhat self-contained and systematic presentation, as well as to carry out various simplifications of detail suggested by a comparative study of the literature. In particular, *the theory of the Lebesgue area of surfaces will not be presupposed. On the contrary, we shall find that the most advanced results of that theory are simple corollaries of our results on general double integrals.*

CHAPTER 1. PRELIMINARIES(2)

1.1. We shall be concerned with integrands of the form $f(x^1, x^2, x^3, X^1, X^2, X^3)$, defined for all values of the six independent variables $x^1, x^2, x^3, X^1, X^2, X^3$. We shall follow the condensed notations used in McShane [1, 2]. Accordingly, we shall use x to refer to the triple x^1, x^2, x^3 and X to refer to the triple X^1, X^2, X^3. It will be convenient for us to interpret x as a point with coordinates x^1, x^2, x^3 and X as a vector with components X^1, X^2, X^3. We shall use $\|X\|$ to denote the length of the vector X. If $X\neq 0$, then ${}_uX$ will denote the unit vector $X/\|X\|$. We shall write $f(x, X)$ for $f(x^1, x^2, x^3, X^1, X^2, X^3)$.

Presented to the Society, April 11, 1941, under the title *On the semi-continuity of double integrals in the calculus of variations*; received by the editors February 3, 1941.

(1) See McShane [1, 2]. Numbers in square brackets refer to the references at the end of this paper.

(2) For the convenience of the reader, we have collected in this chapter the notations, definitions and lemmas we need. *The reader is requested to turn to this chapter whenever in doubt about the meaning of a term or symbol.*

1.2. We shall say that the integrand f is *admissible* if the following conditions are satisfied.

(1) f is continuous for all values of the six independent variables.

(2) f is positively homogeneous of degree one with respect to X. That is,

$$f(x, tX) = tf(x, X) \qquad \text{for } t \geqq 0.$$

(3) f has continuous partial derivatives of the first and second order for $X \neq 0$.

REMARK. As a consequence of (2), we have

$$f(x^1, x^2, x^3, 0, 0, 0) = f(x, 0) = 0.$$

1.3. Let us put

$$f_i = \begin{cases} \dfrac{\partial f}{\partial X_i} & \text{if } X \neq 0, \\ 0 & \text{if } X = 0, \end{cases} \qquad i = 1, 2, 3.$$

We have then, by (2) in 1.2, the identity(3)

$$f(x, X) = X^\alpha f_\alpha(x, X).$$

We define, as usual,

$$E(x^1, x^2, x^3, X^1, X^2, X^3, \overline{X}^1, \overline{X}^2, \overline{X}^3) = E(x, X, \overline{X}) = f(x, \overline{X}) - \overline{X}^\alpha f_\alpha(x, X).$$

1.4. For an admissible f the following facts are easily established. If A is a bounded closed set in x-space, then there exists a constant $M>0$ such that, for $x \in A$ and for every X

$$| f(x, X) | \leqq M\|X\|;$$

and, for every $x \in A$ and for every pair of vectors $X \neq 0$, $\overline{X} \neq 0$,

$$| E(x, X, \overline{X}) | \leqq M\|\overline{X}\| \cdot \|{}_u\overline{X} - {}_uX\|^2 (4).$$

1.5. LEMMA. *Let there be given an admissible f and a set of six constants $(x_0^1, x_0^2, x_0^3, X_0^1, X_0^2, X_0^3) = (x_0, X_0)$, such that*

(a) $f(x_0, X_0) > 0$,

(b) $E(x_0, X_0, \overline{X}) > 0$ *whenever* $\overline{X} \neq 0$, ${}_u\overline{X} \neq {}_uX_0$(5).

Then there exist two positive constants δ_1, δ_2 such that the following holds. If η, $\bar{x}^1$, $\bar{x}^2$, $\bar{x}^3$, $\overline{X}^1$, $\overline{X}^2$, $\overline{X}^3$ satisfy the conditions

(α) $0 < \eta < \delta_1$,

(β) $\overline{X}^\alpha f_\alpha(x_0, X_0) > 0$,

(γ) $|f(\bar{x}, \overline{X}) - f(x_0, \overline{X})| \leqq \eta\|\overline{X}\|$,

---

(3) A repeated greek letter indicates summation with respect to that letter.

(4) Cf. Bliss [1].

(5) Observe that $X_0 \neq 0$ on account of condition (a).

then

$$(1) \qquad f(\bar{x}, \overline{X}) - f(x_0, \overline{X}) + E(x_0, X_0, \overline{X}) \geqq -\eta\delta_2 \overline{X}^\alpha f_\alpha(x_0, X_0).$$

Proof. By condition (a), we have $X_0 \neq 0$, and hence we can consider ${}_uX_0$. For $\sigma > 0$, let us denote by $\lambda(\sigma)$ the maximum of

$$\| {}_u\overline{X} - {}_uX_0 \|$$

for all vectors $\overline{X} \neq 0$ such that

$$E(x_0, X_0, \overline{X}) \leqq \sigma \| \overline{X} \|.$$

From condition (b) we infer easily that $\lambda(\sigma) \to 0$ for $\sigma \to 0$. Hence (cf. condition (a)), we have a $\delta_1 > 0$ such that

$$\lambda(\delta_1) < \frac{f(x_0, X_0)}{2\|X_0\| \cdot \|Y_0\|},$$

where Y_0 denotes the vector with components $f_i(x_0, X_0)$, $i = 1, 2, 3$. Note that condition (a) implies that $Y_0 \neq 0$. We define now

$$\delta_2 = \frac{2\|X_0\|}{f(x_0, X_0)},$$

and we assert that the constants δ_1 and δ_2 satisfy the requirements of the lemma. To show this, let there be given η and $(\bar{x}, \overline{X})$ such that conditions (α), (β), (γ) are satisfied. Let us put

$$H = f(\bar{x}, \overline{X}) - f(x_0, \overline{X}) + E(x_0, X_0, \overline{X}).$$

If $H \geqq 0$, then (1) is true on account of (β). Thus we can assume that $H < 0$. We have then, on account of (γ) and (α),

$$(2) \qquad E(x_0, X_0, \overline{X}) < f(x_0, \overline{X}) - f(\bar{x}, \overline{X}) \leqq \eta \|\overline{X}\| \leqq \delta_1 \|\overline{X}\|.$$

We note that $\overline{X} \neq 0$ on account of (β). Hence we can consider ${}_u\overline{X}$. By the definition of δ_1, (2) implies that

$$(3) \qquad \| {}_u\overline{X} - {}_uX_0 \| \leqq \lambda(\delta_1) < \frac{f(x_0, X_0)}{2\|X_0\| \cdot \|Y_0\|}.$$

The inequality of Schwarz yields

$$(4) \qquad | {}_u\overline{X}^\alpha f_\alpha(x_0, X_0) - {}_uX_0^\alpha f_\alpha(x_0, X_0) | \leqq \|Y_0\| \cdot \| {}_u\overline{X} - {}_uX_0 \|;$$

(4) and (3) yield

$$| {}_u\overline{X}^\alpha f_\alpha(x_0, X_0) - {}_uX_0^\alpha f_\alpha(x_0, X_0) | < \frac{f(x_0, X_0)}{2\|X_0\|} = \frac{1}{\delta_2}.$$

Hence

$$ {}_u\overline{X}^{\alpha} f_\alpha(x_0, X_0) > {}_uX_0^{\alpha} f_\alpha(x_0, X_0) - \frac{1}{\delta_2} = \frac{f(x_0, X_0)}{\|X_0\|} - \frac{1}{\delta_2} = \frac{1}{\delta_2}. $$

Multiplication by $\|X\|$ yields

$$ \overline{X}^{\alpha} f_\alpha(x_0, X_0) > \frac{1}{\delta_2}\|\overline{X}\|; \tag{5} $$

(5) and conditions (b), (γ) yield finally

$$ H > -\eta\|\overline{X}\| > -\eta\delta_2\overline{X}^{\alpha} f_\alpha(x_0, X_0), $$

and the lemma is proved.

1.6. We shall consider triples of functions $x^i(u^1, u^2)$, $i=1, 2, 3$, where the range of definition will be some simply connected Jordan region B, that is, the set of points in and on some Jordan curve. We shall use the notations

$$ T:\quad x^1(u^1, u^2),\ x^2(u^1, u^2),\ x^3(u^1, u^2), \qquad (u^1, u^2) \in B, $$

or in condensed form,

$$ T:\quad x(u), \qquad u \in B, $$

or simply (T, B) to refer to such a triple. The set of points, in (x^1, x^2, x^3)-space, with coordinates $x^1(u^1, u^2)$, $x^2(u^1, u^2)$, $x^3(u^1, u^2)$, will be denoted by $\sum(T, B)$.

1.7. The triple (T, B) will be called *quasi-linear* if the following conditions are satisfied.

(a) The boundary of B is a *polygon*.

(b) $x^1(u^1, u^2)$, $x^2(u^1, u^2)$, $x^3(u^1, u^2)$ are continuous in B.

(c) B can be subdivided into a finite number of (rectilinear) triangles in each of which the functions $x^i(u_1, u_2)$, $i=1, 2, 3$, are linear.

1.8. We shall say that the triple (T, B) is of class K_1 if the functions $x^1(u^1, u^2)$, $x^2(u^1, u^2)$, $x^3(u^1, u^2)$ are continuous in B, if their partial derivatives of the first order exist a.e. in B^0, and if the Jacobians

$$ X^1(u^1, u^2) = \frac{\partial(x^2, x^3)}{\partial(u^1, u^2)}, \quad X^2(u^1, u^2) = \frac{\partial(x^3, x^1)}{\partial(u^1, u^2)}, \quad X^3(u^1, u^2) = \frac{\partial(x^1, x^2)}{\partial(u^1, u^2)} $$

are summable in B^0[6]. We shall denote by $X(u)$ the vector whose components are $X^1(u)$, $X^2(u)$, $X^3(u)$. If $(T, B) \in K_1$, then clearly $\|X(u)\|$ is summable in B^0.

1.9. If $(T, B) \in K_1$ and if f is admissible, then clearly $f(x(u), X(u))$ is measurable in B^0. Since the set $\sum(T, B)$ is bounded and closed, we have by 1.4 a constant $M>0$ such that

$$ |f(x(u), X(u))| \leq M\|X(u)\| \quad \text{a.e. in } B^0. $$

(6) Generally, if H is a point set considered in some space, then H^0 will denote the set of its interior points relative to that space.

Hence $f(x(u), X(u))$ is summable in B^0. We shall denote its integral over B^0 by $I(T, B, f)$. For example,

$$I(T, B, \|X\|) = \iint_{B^0} \|X(u)\| du(^7).$$

1.10. Let there be given a quasi-linear triple

$$T: \quad x(u), \qquad u \in B.$$

We subject B to a topological transformation $\bar{u} = \tau(u)$ and for clarity we want to think of the image point $\bar{u}$ as being located in a different plane $(\bar{u}^1, \bar{u}^2)$. Let us denote by $\overline{B}$ the image of B, and by $u = \sigma(\bar{u})$ the inverse of τ. We have then the new triple

$$\overline{T}: \quad \bar{x}(\bar{u}) = x(\sigma(\bar{u})), \qquad \bar{u} \in \overline{B}.$$

Suppose now that τ is quasi-linear (that is, B can be subdivided into a finite number of rectilinear triangles in each of which τ is an affine transformation). Then clearly $(\overline{T}, \overline{B})$ is quasi-linear. Given then an admissible f, we find by an entirely elementary computation the formula

$$I(T, B, f) = I(\overline{T}, \overline{B}, f),$$

if τ is sense-preserving, and the formula

$$I(T, B, f) = I(\overline{T}, \overline{B}, f^*),$$

where $f^*(x, X) = f(x, -X)$, if τ is not sense-preserving.

1.11. Let there be given an admissible f and a triple

$$T_0: \quad x_0(u), \qquad u \in B_0$$

of class K_1. We shall say that (T_0, B_0) satisfies condition (c) with respect to f if

(1) there exists, in x-space, a closed bounded set A such that $\sum(T_0, B_0) \subset A^0$, and $f(x, X) \geqq 0$ for $x \in A$ and for every vector X, and

(2) for a.e. point $u \in B_0^0$ such that $X_0(u)$ exists and is not equal to 0, we have

$$E(x_0(u), X_0(u), \overline{X}) \geqq 0$$

for every vector $\overline{X} \neq 0$(6).

1.12. We shall say that (T_0, B_0) satisfies condition (+c) with respect to f if condition (2) in 1.11 is satisfied, while condition (1) there is satisfied in the following stronger form:

There exists, in x-space, a closed bounded set A such that $\sum(T_0, B_0) \subset A^0$, and $f(x, X) > 0$ for every $x \in A$ and for every vector $X \neq 0$.

1.13. We shall say that (T_0, B_0) satisfies condition (c+) with respect to f if condition (1) in 1.11 is satisfied, while condition (2) there is satisfied in the following stronger form:

---

(7) For brevity we write du for du^1du^2.

For a.e. point $u \in B_0^0$ such that $X_0(u)$ exists and is not equal to 0, we have

$$E(x_0(u), X_0(u), \overline{X}) > 0$$

for every vector $\overline{X} \neq 0$ such that ${}_u\overline{X} \neq {}_uX_0(u)$.

1.14. For the particular integrand $\|X\|$ we find

$$E(X, \overline{X}) = \tfrac{1}{2}\|\overline{X}\| \cdot \|{}_u\overline{X} - {}_uX\|^2.$$

This identity leads immediately to the following lemmas.

1.15. Lemma. *If f is admissible and if $(T_0, B_0) \in K_1$, then for a sufficiently large value of the constant H the triple (T_0, B_0) will satisfy condition* (c+) *with respect to the integrand $H\|X\|+f$.*

1.16. Lemma. *If f is admissible, if $(T_0, B_0) \in K_1$, and if (T_0, B_0) satisfies condition* (c) *with respect to f, then for every $\epsilon > 0$ the triple (T_0, B_0) satisfies condition* (c+) *with respect to the integrand $\epsilon\|X\|+f$.*

1.17. Let there be given two continuous triples

$$T: \; x(u), \quad u \in B; \qquad \overline{T}: \; \bar{x}(\bar{u}), \quad \bar{u} \in \overline{B},$$

where, for clarity, we want to think of $\overline{B}$ as being located in a different plane $\bar{u}$. Let $\bar{u} = \tau(u)$ be a topological transformation from B to $\overline{B}$. The *Fréchet distance* of (T, B) and $(\overline{T}, \overline{B})$ is then defined as the greatest lower bound, for all possible topological transformations τ, of

$$\max_{u \in B} \|x(u) - \bar{x}(\tau(u))\| \text{(8)}.$$

We shall denote this distance by $d[(T, B), (\overline{T}, \overline{B})]$.

1.18. If in the preceding definition we restrict τ to be *sense-preserving*, then we obtain what may be called *the oriented distance* of (T, B) and $(\overline{T}, \overline{B})$. We shall denote the oriented distance by ${}_od[(T, B), (\overline{T}, \overline{B})]$(9).

1.19. We shall say that a triple (T_0, B_0) is of class K_2 if the following conditions are satisfied(10).

(a) $(T_0, B_0) \in K_1$.

(b) There exists a sequence of quasi-linear triples (T_n, B_n) such that

$${}_od[(T_0, B_0), (T_n, B_n)] \xrightarrow[n\to\infty]{} 0,$$

and

$$I(T_n, B_n, \|X\|) \xrightarrow[n\to\infty]{} I(T_0, B_0, \|X\|).$$

1.20. Given a continuous triple (T_0, B_0), let us consider the class $\mathfrak{K}$ of all

---

(8) If $p, \bar{p}$ are points in a metric space, then $\|p - \bar{p}\|$ will denote their distance.

(9) This modification of the Fréchet distance was introduced by McShane, loc. cit.(1).

(10) As we shall prove later on, *these conditions are necessary and sufficient that the Lebesgue area of the surface determined by (T_0, B_0) be given by the usual integral formula.*

continuous triples (T, B) such that

$$d[(T_0, B_0), (T, B)] = 0.$$

Clearly, $\mathfrak{K}$ is univocally determined by any one of its elements, and possesses the following properties:

(a) If $(T_1, B_1) \in \mathfrak{K}$, $(T_2, B_2) \in \mathfrak{K}$, then $d[(T_1, B_1), (T_2, B_2)] = 0$.

(b) If $(T_1, B_1) \in \mathfrak{K}$ and $d[(T_1, B_1), (T_2, B_2)] = 0$, then $(T_2, B_2) \in \mathfrak{K}$.

Conversely, every class of triples, with the properties (a) and (b), can be generated by any one of its elements in the manner described above.

1.21. A *continuous parametrized surface* S, *of the type of the circular disc*, is merely a class $\mathfrak{K}$ of continuous triples as described in 1.20. Every triple $(T, B) \in \mathfrak{K}$ will be called a representation of S. For brevity, we shall call S simply a *continuous surface.*

1.22. Given two continuous surfaces S_1, S_2, let (T_1, B_1) be a representation of S_1 and (T_2, B_2) a representation of S_2. Then $d[(T_1, B_1), (T_2, B_2)]$ is easily seen to be independent of the particular choice of these representations and can be denoted therefore by $d(S_1, S_2)$. This quantity is *the Fréchet distance of the surfaces* S_1 *and* S_2.

If (T_1, B_1), (T_2, B_2) are any two representations of the same continuous surface S, then the point sets $\sum(T_1, B_1)$, $\sum(T_2, B_2)$ are easily seen to be identical. That is, the set $\sum(T, B)$ is the same for all representations of (T, B). We shall denote this set by $\sum(S)$.

1.23. If in the statements made in 1.20, 1.21, 1.22 we replace the Fréchet distance d by the *oriented* distance ${}_od$, then we arrive at the conception of an *oriented continuous parametrized surface, of the type of the circular disc.* We shall denote such a surface by ${}_oS$ and we shall call it simply an *oriented continuous surface.* The definition of the oriented distance ${}_od({}_oS_1, {}_oS_2)$ of two oriented continuous surfaces ${}_oS_1$, ${}_oS_2$ as well as the definition of the point set $\sum({}_oS)$ associated with an oriented continuous surface ${}_oS$ is then worded in an obvious manner (cf. 1.22).

1.24. If S_0, S_n are continuous surfaces, then $S_n \to S_0$ will mean that $d(S_n, S_0) \to 0$. If $S_n \to S_0$, and if A is a point set in (x^1, x^2, x^3)-space such that $\sum(S_0) \subset A^0$(11), then clearly $\sum(S_n) \subset A^0$ for large values of n.

1.25. If ${}_oS_0$, ${}_oS_n$ are oriented continuous surfaces, then ${}_oS_n \to {}_oS_0$ means that ${}_od({}_oS_0, {}_oS_n) \to 0$. That is, for *oriented* continuous surfaces the notion of convergence is based on the *oriented* distance. Again, if ${}_oS_n \to {}_oS_0$, and if A is a point set in (x^1, x^2, x^3)-space such that $\sum({}_oS_0) \subset A^0$, then $\sum({}_oS_n) \subset A^0$ for large values of n.

1.26. We say that a continuous surface, oriented or not, is of class K_i if it admits of a representation (T, B) that is of class K_i, $i = 1, 2$ (cf. 1.8, 1.19).

1.27. In the u-plane, let there be given a simply connected bounded Jordan region $\overline{B}$. In a $\bar{u}$-plane let there be given a region $\overline{\bar{B}}$ of the same type.

(11) Cf. (8).

We assume that $\bar{B}$ is bounded by a polygon. Let there be given also a topological transformation $u=\tau(\bar{u})$ from $\bar{B}$ to B. Given then $\epsilon>0$, there exists a quasi-linear topological transformation $u=\tau^*(\bar{u})$ from $\bar{B}$ to some Jordan region B^* in the u-plane, such that

$$\|\tau(\bar{u}) - \tau^*(\bar{u})\| \leqq \epsilon \qquad \text{for } \bar{u} \in \bar{B}(^{12}).$$

If τ was sense-preserving, then τ^* can be chosen as sense-preserving (in fact, for ϵ small, τ^* will be then automatically sense-preserving)([13]).

1.28. GENERALIZATION OF A LEMMA OF MCSHANE. *In a closed square Q, in the u-plane, let there be given triples*

$$T_0:\ x_0(u); \quad u \in Q, \qquad T_n:\ x_n(u), \quad u \in Q,$$

such that the following conditions are satisfied.

(a) $(T_0, Q) \in K_1$.

(b) (T_n, Q) *is quasi-linear.*

(c) $x_n(u) \to x_0(u)$ *uniformly in* Q.

Then for every choice of the constants a^1, a^2, a^3 there exists a sequence of measurable sets V_n in Q, such that

$$\iint_{V_n} a^\alpha X_n^\alpha(u)du \to \iint_Q a^\alpha X_0^\alpha(u)du.$$

COROLLARY. *If*

$$\iint_Q a^\alpha X_0^\alpha(u)du > 0,$$

then the sets V_n can be chosen in such a way that $a^\alpha X_n^\alpha(u) > 0$ on V_n.

1.29. McShane([14]) proved this lemma under the following additional assumptions concerning the limit triple (T_0, Q).

(d) $x_0(u)$ is absolutely continuous on the perimeter p of Q.

(e) The following equations hold

$$\iint_Q X_0^1(u)du = \frac{1}{2}\int_p (x_0^2(u)dx_0^3(u) - x_0^3(u)dx_0^2(u)),$$

$$\iint_Q X_0^2(u)du = \frac{1}{2}\int_p (x_0^3(u)dx_0^1(u) - x_0^1(u)dx_0^3(u)),$$

$$\iint_Q X_0^3(u)du = \frac{1}{2}\int_p (x_0^1(u)dx_0^2(u) - x_0^2(u)dx_0^1(u)).$$

([12]) Cf. ([9]).

([13]) The existence of such quasi-linear approximations follows from Franklin-Wiener [1].

([14]) Loc. cit. ([1]).

It is fundamental for our purposes that these additional assumptions are unnecessary(16). It is interesting to note that *we require solely summability of the Jacobians of the limit triple, a requirement which is necessary if the lemma is to have a meaning.*

1.30. Since (loc. cit.(15)) we did not state the corollary to the lemma, let us suggest briefly how the corollary may be derived from the lemma itself. Let us denote by Q_r the square with the same center as Q, with sides parallel to those of Q, and with side-length equal to r times the side-length of Q, $0<r\leqq 1$. Denote by V_n^+ the subset of V_n where

$$a^{\alpha}X_n^{\alpha}(u) > 0.$$

We define now a measurable subset W_n of V_n as follows. If

$$\iint_{V_n^+} a^{\alpha}X_n^{\alpha}(u)du \leqq \iint_Q a^{\alpha}X_0^{\alpha}(u)du,$$

then $W=V_n^+$. If

$$\iint_{V_n^+} a^{\alpha}X_n^{\alpha}(u)du > \iint_Q a^{\alpha}X_0^{\alpha}(u)du, \tag{6}$$

then let us denote by $E_{n,r}$ the set $Q_r\cdot V_n^+$. Then the quantity

$$\mu_n(r) = \iint_{E_{n,r}} a^{\alpha}X_n^{\alpha}(u)du \tag{7}$$

is a continuous function of r for $0<r\leqq 1$. Clearly $\mu_n(r)\rightarrow_{r\rightarrow 0}0$ and, by (6), (7),

$$\mu_n(1) > \iint_Q a^{\alpha}X_0^{\alpha}(u)du.$$

Hence there exists a value r_n between zero and one such that

$$\mu_n(r_n) = \iint_{E_{n,r_n}} a^{\alpha}X_n^{\alpha}(u)du = \iint_Q a^{\alpha}X_0^{\alpha}(u)du.$$

We put then $W_n=E_{n,r_n}$. For the sequence of sets W_n defined in this manner we have obviously

$$\iint_{W_n} a^{\alpha}X_n^{\alpha}(u)du \rightarrow \iint_Q a^{\alpha}X_0^{\alpha}(u)du,$$

and

$$a^{\alpha}X_n^{\alpha}(u) > 0 \quad \text{on} \quad W_n.$$

(16) See Radó [4].

1.31. **On functions of squares**(16). In the (u^1, u^2)-plane, we consider a fixed square Q_0 and all squares $q \subset Q_0$, all these squares having their sides parallel to the axes u^1, u^2. Given then a function $\psi(q)$ for all such squares, including Q_0 itself(17), its upper derivative $\overline{D}\psi(u)$ at an interior point u of Q_0 is defined as the least upper bound of

$$\limsup_{n\to\infty} \frac{\psi(q_n)}{|q_n|}$$

for all possible sequences of squares q_n, with sides parallel to the axes, containing u, and such that $|q_n| \to 0$. The lower derivative $\underline{D}\psi(u)$ is defined in a similar fashion. Both of these derivatives are measurable functions. Generally, they will take on the values $\pm\infty$. If $\overline{D}\psi(u)$ and $\underline{D}\psi(u)$ are finite and equal at a point u, then their common value is the derivative $D\psi(u)$. The set on which $D\psi(u)$ exists is measurable and $D\psi(u)$ is measurable on this set. In particular if $D\psi(u)$ exists a.e. in Q_0, then it is measurable in Q_0.

1.32. Let $q, q_1, q_2, \cdots, q_m$ be a system of squares in Q_0 such that $q_i \subset q$ for $i=1, 2, \cdots, m$ and such that $q_i^0 \cdot q_j^0 = 0$ for $i \neq j$. If for every such system we have

$$\sum_1^m \psi(q_i) \leqq \psi(q),$$

if $\psi(q) \geqq 0$ for every $q \subset Q_0$, and if $\psi(Q_0) < +\infty$, then we shall say that $\psi(q)$ is of type A in Q_0.

1.33. Lemma. *If $\psi(q)$ is of type A in Q_0, then its derivative $D\psi(u)$ exists a.e. in Q_0, is summable there, and satisfies the inequality*

$$\iint_q D\psi(u)du \leqq \psi(q)$$

for every $q \subset Q_0$(18).

1.34. The fundamental pattern of our semi-continuity proofs may be now described by the following

Lemma. *In the u-plane, let there be given bounded and simply connected Jordan regions $B_0, B_1, B_2, \cdots, B_n, \cdots$ such that the following condition holds: If q is any closed square in B_0^0*(19)*, then $q \subset B_n^0$ for sufficiently large values of n. In B_n^0 let there be given a non-negative summable function $f_n(u)$, $n=0, 1, 2, \cdots$. No assumption is made concerning the convergence of the sequence $f_n(u)$. For*

(16) The facts stated in 1.31, 1.32, 1.33 are immediate consequences of results in Banach [1]. Cf. also Radó [3].

(17) It is convenient to permit $\psi(q)$ to assume infinite values.

(18) While this lemma is implied by the results in Banach [1], it seems that it was first stated and used explicitly in Radó [2].

(19) Cf. (8).

every closed square $q \subset B_0^0$ let us define

$$\psi(q) = \liminf_{n\to\infty} \iint_q f_n(u)du. \tag{8}$$

If we have, for a.e. point u in B_0^0, the inequality

$$\underline{D}\psi(u) \geqq f_0(u), \tag{9}$$

then

$$\liminf_{n\to\infty} \iint_{B_n^0} f_n(u)du \geqq \iint_{B_0^0} f_0(u)du. \tag{10}$$

Proof[20]. Let Q be a fixed closed square in B_0^0. Since $f_n(u) \geqq 0$, it is obvious that $\psi(q)$ is of type A in Q. Hence, by 1.33,

$$\psi(Q) \geqq \iint_Q \underline{D}\psi(u)du. \tag{11}$$

Relations (8), (9), (11) yield

$$\liminf \iint_Q f_n(u)du \geqq \iint_Q f_0(u)du \tag{12}$$

for every $Q \subset B_0^0$. Given now any $\epsilon > 0$, we can select in B_0^0 a finite number of closed squares $Q_1, Q_2, \cdots, Q_m$ without common interior points, such that

$$\sum_1^m \iint_{Q_i} f_0(u)du > \iint_{B_0^0} f_0(u)du - \epsilon.$$

We have an N such that $Q_1, Q_2, \cdots, Q_m \subset B_n^0$ for $n > N$. Since $f_n \geqq 0$, we have then for $n > N$

$$\iint_{B_n^0} f_n(u)du \geqq \sum_1^m \iint_{Q_i} f_n(u)du,$$

and hence, by (12),

$$\begin{aligned}\liminf_{n\to\infty} \iint_{B_n^0} f_n(u)du &\geqq \sum_1^m \liminf_{n\to\infty} \iint_{Q_i} f_n(u)du \\ &\geqq \sum_1^m \iint_{Q_i} f_0(u)du > \iint_{B_0^0} f_0(u)du - \epsilon.\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, (10) is proved.

---

[20] We assume that $\psi(q) < +\infty$ for every $q \in B_0^0$. This assumption is justified by the remark that if $\psi(q) = +\infty$ for some $q \subset B_0^0$, then clearly the left-hand member in (10) is $+\infty$ and then (10) is obvious.

CHAPTER 2. LEMMAS ON TRIPLES

2.1. FUNDAMENTAL LEMMA. *Let there be given an admissible integrand f and triples*

$$T_0:\ x_0(u), \quad u \in B_0, \qquad T_n:\ x_n(u), \quad u \in B_n,$$

such that the following conditions are satisfied.

(1) $(T_0, B_0) \in K_1$.

(2) (T_n, B_n) *is quasi-linear.*

(3) (T_0, B_0) *satisfies condition* (c+) *relative to* f.

(4) *For every closed square* $q \subset B_0^0$ *there exists an* $N = N(q)$ *such that* $q \subset B_n^0$ *for* $n > N$.

(5) *On every closed square* $q \subset B_0^0$, $x_n(u) \rightarrow x_0(u)$ *uniformly.*

(6) $\mathcal{d}[(T_0, B_0), (T_n, B_n)] \rightarrow 0$.

Then

$$\liminf_{n\to\infty} I(T_n, B_n, f) \geqq I(T_0, B_0, f).$$

REMARK. By conditions (6) and (3) we have, for large n, $f(x_n(u), X_n(u)) \geqq 0$ a.e. in B_n^0, and hence the application of 1.34 is justified.

2.2. **Proof.** According to the scheme outlined in 1.34, we define, for every closed square $q \subset B_0^0$,

$$\psi(q) = \liminf I(T_n, q, f).$$

Let u_0 be a point in B_0^0 and q_0 a small square in B_0^0 with center at u_0. Since we wish to obtain information about $\underline{D}\psi(u_0)$ for *almost every* point in B_0^0, we can assume that the following statements hold true for u_0 and q_0. First, $X_0(u_0)$ exists and

$$\frac{\iint_{q_0} X_0^i(u)du}{|q_0|} \rightarrow X_0^i(u_0), \qquad i = 1, 2, 3, \tag{13}$$

if $|q_0| \rightarrow 0$(21). Next, we can assume that (cf. 1.3)

$$f(x_0(u_0), X_0(u_0)) = X_0^\alpha(u_0) f_\alpha(x_0(u_0), X_0(u_0)) > 0. \tag{14}$$

Indeed, we wish to prove that

$$\underline{D}\psi(u_0) \geqq f(x_0(u_0), X_0(u_0)). \tag{15}$$

But $\underline{D}\psi \geqq 0$ and $f(x_0(u_0), X_0(u_0)) \geqq 0$. Thus (15) is obvious if $f(x_0(u_0), X_0(u_0)) = 0$. Now it follows from (13) that

$$\frac{\iint_{q_0} X_0^\alpha(u) f_\alpha(x_0(u_0), X_0(u_0))du}{|q_0|} \rightarrow X_0^\alpha(u_0) f_\alpha(x_0(u_0), X_0(u_0)) \tag{16}$$

(21) If H is a measurable point set, then $|H|$ denotes its measure.

if $|q_0| \to 0$. By restricting the size of q_0, we shall have therefore (cf. (14))

$$\iint_{q_0} X_0^{\alpha}(u) f_{\alpha}(x_0(u_0), X_0(u_0)) du > 0. \tag{17}$$

From (14) it follows that

$$X_0(u_0) \neq 0.$$

Finally, we can assume that $E(x_0(u_0), X_0(u_0), \overline{X}) > 0$ whenever $\overline{X} \neq 0$ and ${}_u\overline{X} \neq {}_u X_0(u_0)$.

2.3. We now apply the lemma of 1.5 to the set of constants $(x_0(u_0), X_0(u_0))$. By that lemma, we have two positive constants δ_1, δ_2 such that whenever $0 < \eta < \delta_1$, we have

$$f(\bar{x}, \overline{X}) - f(x_0(u_0), \overline{X}) + E(x_0(u_0), X_0(u_0), \overline{X}) \geqq -\eta\delta_2 \overline{X}^{\alpha} f_{\alpha}(x_0(u_0), X_0(u_0))$$

for every set $(\bar{x}, \overline{X})$ for which

$$\overline{X}^{\alpha} f_{\alpha}(x_0(u_0), X_0(u_0)) > 0, \qquad |f(\bar{x}, \overline{X}) - f(x_0(u_0), \overline{X})| \leqq \eta \|\overline{X}\|.$$

2.4. We take an η satisfying $0 < \eta < \delta_1$. Let Q_0 be a fixed closed square in B_0^0, with center at u_0. On Q_0, by assumption, $x_n(u) \to x_0(u)$ uniformly. Hence we have an n_0 such that

$$|f(x_n(u), X_n(u)) - f(x_0(u), X_n(u))| \leqq (\eta/2)\|X_n(u)\| \text{ for } n > n_0, \text{ a.e. } u \in Q_0. \tag{18}$$

If we make the square q_0 of 2.2 sufficiently small, we shall have

$$q_0 \subset Q_0;$$

and, since $x_0(u)$ is continuous,

$$|f(x_0(u), X_n(u)) - f(x_0(u_0), X_n(u))| \leqq (\eta/2)\|X_n(u)\| \quad \text{for a.e. } u \in q_0. \tag{19}$$

Combining (18), (19) we see that

$$|f(x_n(u), X_n(u)) - f(x_0(u_0), X_n(u))| \leqq \eta\|X_n(u)\| \quad \text{for a.e. } u \in q_0,\ n > n_0. \tag{20}$$

By condition (c+) we have an $m_0 = m_0(q_0)$ such that

$$f(x_n(u), X_n(u)) \geqq 0 \qquad \text{for a.e. } u \in q_0,\ n > m_0. \tag{21}$$

We take now first a q_0 so small and then an n so large that all the relations (17), (20), (21) hold.

2.5. By 1.28 we have in q_0 a sequence of measurable sets V_n such that

$$\iint_{V_n} X_n^{\alpha}(u) f_{\alpha}(x_0(u_0), X_0(u_0)) du \xrightarrow[n \to \infty]{} \iint_{q_0} X_0^{\alpha}(u) f_{\alpha}(x_0(u_0), X_0(u_0)) du, \tag{22}$$

and (cf. (17) and 1.28, corollary)

$$X_n^{\alpha}(u) f_{\alpha}(x_0(u_0), X_0(u_0)) > 0 \qquad \text{for } u \in V_n. \tag{23}$$

We define

$$\overline{X}_n(u) = \begin{cases} X_n(u) & \text{on} \quad V_n, \\ 0 & \text{on} \quad q_0 - V_n. \end{cases} \tag{24}$$

By (21), (24), 1.3, 2.3, (23), (24) we can write now, for a.e. $u \in q_0$,

$$\begin{aligned} f(x_n(u), X_n(u)) &\geqq f(x_n(u), \overline{X}_n(u)) \\ &= f(x_n(u), \overline{X}_n(u)) - f(x_0(u_0), \overline{X}_n(u)) + E(x_0(u_0), X_0(u_0), \overline{X}_n(u)) \\ &\quad + \overline{X}_n^\alpha(u) f_\alpha(x_0(u_0), X_0(u_0)) \\ &\geqq (1 - \eta\delta_2)\overline{X}_n^\alpha(u) f_\alpha(x_0(u_0), X_0(u_0))du. \end{aligned}$$

Integrate over q_0 and let $n \to \infty$. By (22), (24) it follows that

$$\psi(q_0) \geqq (1 - \eta\delta_2) \iint_{q_0} X_0^\alpha(u) f_\alpha(x_0(u_0), X_0(u_0))du.$$

Divide by $|q_0|$ and let $|q_0| \to 0$. By (16) it follows that

$$\underline{D}\psi(u_0) \geqq (1 - \eta\delta_2)X_0^\alpha(u_0) f_\alpha(x_0(u_0), X_0(u_0)) = (1 - \eta\delta_2) f(x_0(u_0), X_0(u_0)).$$

Since η was arbitrary, this shows that

$$\underline{D}\psi(u_0) \geqq f(x_0(u_0), X_0(u_0)).$$

On account of 1.34, this proves the lemma of 2.1.

2.6. Lemma. *Let there be given an admissible integrand f and triples (T_0, B_0), (T_n, B_n), such that the following conditions are satisfied.*

(1) $(T_0, B_0) \in K_1$.

(2) (T_n, B_n) *is quasi-linear.*

(3) $\mathcal{d}[(T_0, B_0), (T_n, B_n)] \to 0$.

(4) (T_0, B_0) *satisfies condition* (c+) *relative to f.*

Then

$$\liminf I(T_n, B_n, f) \geqq I(T_0, B_0, f). \tag{25}$$

2.7. **Proof.** For clarity, we want to think of B_n as being located in a different plane $\bar{u}$. We write therefore explicitly

$$T_0:\ x_0(u), \quad u \in B_0, \qquad T_n:\ x_n(\bar{u}), \quad \bar{u} \in B_n.$$

Let us put

$$\delta_n = \mathcal{d}[(T_0, B_0), (T_n, B_n)].$$

We have then, for every n, a sense-preserving topological transformation

$$\bar{u} = \tau_n(u), \qquad u \in B_0,$$

which carries B_0 into B_n and for which

$$\|x_0(u) - x_n(\tau_n(u))\| < \delta_n + 1/n. \tag{26}$$

Let

$$u = \sigma_n(\bar{u}), \qquad \bar{u} \in B_n,$$

be the inverse of τ_n. Since B_n is a polygon, we have (cf. 1.27) a sense-preserving, topological, quasi-linear transformation

$$u = \sigma_n^*(\bar{u}), \qquad \bar{u} \in B_n,$$

which carries B_n into some polygonal region B_n^* in the u-plane (where B_0 itself is located) and which satisfies the condition

$$\|\sigma_n(\bar{u}) - \sigma_n^*(\bar{u})\| < 1/n, \qquad \bar{u} \in B_n\text{[22]}. \tag{27}$$

Let us denote by

$$\bar{u} = \tau_n^*(u), \quad u \in B_n^*,$$

the inverse of σ_n^* and let us introduce the triple

$$T_n^*: \quad x_n(\tau_n^*(u)), \quad u \in B_n^*.$$

Since (T_n, B_n) and τ_n^* are quasi-linear and since τ_n^* is sense-preserving, it follows by 1.10 that

$$I(T_n, B_n, f) = I(T_n^*, B_n^*, f), \tag{28}$$

and further, on account of (26) and (27), that the triples (T_0, B_0), (T_n^*, B_n^*) and the integrand f satisfy the assumptions of the lemma in 2.1. Hence, by that lemma,

$$\liminf I(T_n^*, B_n^*, f) \geqq I(T_0, B_0, f); \tag{29}$$

(28) and (29) imply (25).

2.8. Lemma. *If* $(T_0, B_0) \in K_2$, $(\overline{T}_0, \overline{B}_0) \in K_2$, *and if*

$${}_0d[(T_0, B_0), (\overline{T}_0, \overline{B}_0)] = 0, \tag{30}$$

then

$$I(T_0, B_0, \|X\|) = I(\overline{T}_0, \overline{B}_0, \|X\|). \tag{31}$$

Proof. Since $(T_0, B_0) \in K_2$, we have a sequence of quasi-linear triples (T_n, B_n) such that

$${}_0d[(T_0, B_0), (T_n, B_n)] \to 0, \tag{32}$$

and

---

[22] Cf. [8].

$$I(T_n, B_n, \|X\|) \to I(T_0, B_0, \|X\|); \tag{33}$$

(30) and (32) imply that

$$_0d[(\overline{T}_0, \overline{B}_0), (T_n, B_n)] \to 0.$$

Applying the lemma of 2.6 to the triples $(\overline{T}_0, \overline{B}_0)$, (T_n, B_n) with $f = \|X\|$, we obtain

$$\liminf I(T_n, B_n, \|X\|) \geqq I(\overline{T}_0, \overline{B}_0, \|X\|), \tag{34}$$

and (33) and (34) yield

$$I(T_0, B_0, \|X\|) \geqq I(\overline{T}_0, \overline{B}_0, \|X\|). \tag{35}$$

The complementary inequality

$$I(\overline{T}_0, \overline{B}_0, \|X\|) \geqq I(T_0, B_0, \|X\|) \tag{36}$$

is obtained in the same manner, (35) and (36) imply (31).

2.9. Lemma. *Let there be given an admissible integrand f and triples* (T_0, B_0), (T_n, B_n) *such that the following conditions are satisfied.*

(1) $(T_0, B_0) \in K_2$.

(2) (T_n, B_n) *is quasi-linear.*

(3) $_0d[(T_0, B_0), (T_n, B_n)] \to 0$.

(4) $I(T_n, B_n, \|X\|) \to I(T_0, B_0, \|X\|)$.

Then

$$I(T_n, B_n, f) \to I(T_0, B_0, f). \tag{37}$$

Proof. From 1.15 it follows easily that if the positive constant H is sufficiently large, then (T_0, B_0) satisfies condition (c+) relative to the integrand

$$H\|X\| + f.$$

Hence we have, by the lemma of 2.6,

$$\liminf \{HI(T_n, B_n, \|X\|) + I(T_n, B_n, f)\} \geqq HI(T_0, B_0, \|X\|) + I(T_0, B_0, f).$$

In view of condition (4) of the lemma it follows that

$$\liminf I(T_n, B_n, f) \geqq I(T_0, B_0, f). \tag{38}$$

The same reasoning, applied to the integrand $-f$, yields

$$\liminf (-I(T_n, B_n, f)) \geqq -I(T_0, B_0, f). \tag{39}$$

(38) and (39) imply (37).

2.10. Lemma. *Let there be given an admissible integrand f and triples* (T_0, B_0), $(\overline{T}_0, \overline{B}_0)$ *such that the following conditions are satisfied.*

(1) $(T_0, B_0) \in K_2$.

(2) $(\overline{T}_0, \overline{B}_0) \in K_2$.
(3) ${}_o d[(T_0, B_0), (\overline{T}_0, \overline{B}_0)] = 0$.

Then

$$I(T_0, B_0, f) = I(\overline{T}_0, \overline{B}_0, f). \tag{40}$$

Proof. Since $(\overline{T}_0, \overline{B}_0) \in K_2$, we have a sequence of quasi-linear triples (T_n, B_n) such that

$$ {}_o d[(\overline{T}_0, \overline{B}_0), (T_n, B_n)] \to 0, \tag{41}$$

and

$$I(T_n, B_n, \|X\|) \to I(\overline{T}_0, \overline{B}_0, \|X\|); \tag{42}$$

(41) and the condition (3) above imply that

$$ {}_o d[(T_0, B_0), (T_n, B_n)] \to 0. \tag{43}$$

By 2.8 we have

$$I(\overline{T}_0, \overline{B}_0, \|X\|) = I(T_0, B_0, \|X\|).$$

Hence, by (42),

$$I(T_n, B_n, \|X\|) \to I(T_0, B_0, \|X\|). \tag{44}$$

By 2.9, the relations (41), (42) and (43), (44), respectively, imply that

$$I(T_n, B_n, f) \to I(\overline{T}_0, \overline{B}_0, f) \tag{45}$$

and

$$I(T_n, B_n, f) \to I(T_0, B_0, f). \tag{45*}$$

(45) and (45*) imply (40).

Chapter 3. The principal theorems[23]

3.1. Definition. *A continuous surface of class K_2, oriented or not, satisfies condition* (c), (+c), *or* (c+), *respectively, relative to an admissible integrand f if it possesses a representation of class K_2 that satisfies condition* (c), (+c), *or* (c+), *respectively, relative to f* (cf. 1.11, 1.12, 1.13).

3.2. Theorem. *Let f be an admissible integrand and ${}_oS$ an oriented continuous surface of class K_2. If (T_1, B_1), (T_2, B_2) are any two representations of class K_2 of ${}_oS$, then*

$$I(T_1, B_1, f) = I(T_2, B_2, f).$$

This is a direct consequence of 2.10.

---

(23) Disregarding a few minor items, the semi-continuity theorems of this chapter are generalizations of theorems of McShane. The theorems on the Lebesgue area of surfaces are generalizations of previous results of McShane, Morrey and the author.

3.3. On account of 3.2, if f is admissible and if ${}_oS$ is of class K_2, then $I(T, B, f)$ has the same value for all representations of class K_2 of ${}_oS$. This common value may be denoted therefore by $I({}_oS, f)$.

3.4. We shall denote by ${}_o\mathfrak{C}$ the class of all oriented continuous surfaces of class K_2. By 3.3 the functional $I({}_oS, f)$ is then defined for every admissible f and for every ${}_oS\in{}_o\mathfrak{C}$.

3.5. THEOREM. *For fixed admissible f the functional $I({}_oS, f)$ is lower semi-continuous in ${}_o\mathfrak{C}$ at every oriented continuous surface ${}_oS\in{}_o\mathfrak{C}$ which satisfies condition* (c+) *relative to f.*

Proof. Take any sequence ${}_oS_n$ such that ${}_oS_n\to{}_oS_0$ and ${}_oS_n\in{}_o\mathfrak{C}$. Let (T_0, B_0), (T_n, B_n) be representations of class K_2 of ${}_oS_0$, ${}_oS_n$, respectively, where (T_0, B_0) satisfies condition (c+) relative to f. We have then

$$(46)\qquad {}_od[(T_n, B_n), (T_0, B_0)] \to 0,$$

and by 3.3

$$(47)\qquad I({}_oS_0, f) = I(T_0, B_0, f),$$

$$(48)\qquad I({}_oS_n, f) = I(T_n, B_n, f).$$

Since the triple (T_n, B_n) is of class K_2, we have by definition a sequence of quasi-linear triples (T_n^j, B_n^j) such that

$$(49)\qquad {}_od[(T_n^j, B_n^j), (T_n, B_n)] \xrightarrow[j\to\infty]{} 0,$$

and

$$(50)\qquad I(T_n^j, B_n^j, \|X\|) \xrightarrow[j\to\infty]{} I(T_n, B_n, \|X\|).$$

By 2.9 it follows from (49) and (50) that

$$I(T_n^j, B_n^j, f) \xrightarrow[j\to\infty]{} I(T_n, B_n, f).$$

We have therefore, for every n, a j_n such that

$$(51)\qquad {}_od[(T_n^{j_n}, B_n^{j_n}), (T_n, B_n)] < 1/n,$$

and

$$(52)\qquad I(T_n^{j_n}, B_n^{j_n}, f) < I(T_n, B_n, f) + 1/n.$$

(51) and (46) imply that

$$(53)\qquad {}_od[(T_n^{j_n}, B_n^{j_n}), (T_0, B_0)] \to 0.$$

(53) implies, by 2.6, that

$$\liminf_{n\to\infty} I(T_n^{i_n}, B_n^{i_n}, f) \geqq I(T_0, B_0, f). \tag{54}$$

(47), (48), (52), (54) imply that

$$\liminf I({}_oS_n, f) \geqq I({}_oS_0, f).$$

3.6. THEOREM. *Let f, ${}_oS_0$, ${}_oS_n$ satisfy the following conditions.*

(a) ${}_oS_0 \in {}_o\mathfrak{C}$, ${}_oS_n \in {}_o\mathfrak{C}$.

(b) ${}_oS_n \to {}_oS_0$.

(c) ${}_oS_0$ *satisfies condition* (c) *with respect to the admissible integrand f.*

(d) *There exists a finite constant N such that for every n*

$$I({}_oS_n, \|X\|) < N.$$

Then

$$\liminf I({}_oS_n, f) \geqq I({}_oS_0, f). \tag{55}$$

Proof. Clearly, for every $\epsilon > 0$, ${}_oS_0$ satisfies condition (c+) with respect to the integrand $f + \epsilon\|X\|$ (cf. 1.16). Hence by 3.5

$$\liminf_{n\to\infty} \{I({}_oS_n, f) + \epsilon I({}_oS_n, \|X\|)\} \geqq I({}_oS_0, f) + \epsilon I({}_oS_0, \|X\|).$$

Since $\epsilon > 0$ was arbitrary, (55) follows on account of condition (d).

3.7. THEOREM. *For fixed admissible f, the functional $I({}_oS, f)$ is lower semi-continuous in ${}_o\mathfrak{C}$ at every surface ${}_oS \in {}_o\mathfrak{C}$ that satisfies condition* (+c) *with respect to f.*

Proof. Let us deny the assertion. Then we assert the existence of surfaces ${}_oS_0$, ${}_oS_n$ in ${}_o\mathfrak{C}$ such that ${}_oS_n \to {}_oS_0$ and

$$\lim I({}_oS_n, f) < I({}_oS_0, f), \tag{56}$$

while ${}_oS_0$ satisfies condition (+c) relative to f. This last fact implies the existence, in x-space, of a closed bounded set A such that $\sum({}_oS_0) \subset A^0$ and $f(x, X) > 0$ for $x \in A$ and for every vector $X \neq 0$. It follows that there exists a constant $m > 0$ such that

$$f(x, X) \geqq m\|X\| \qquad \text{for } x \in A \tag{57}$$

and for every vector X. Since ${}_oS_n \to {}_oS_0$, we have an n_0 such that

$$\sum({}_oS_n) \subset A^0 \qquad \text{for } n > n_0. \tag{58}$$

Let

$$T_n: \ x_n(u), \qquad u \in B_n,$$

be a representation of class K_2 of ${}_oS_n$. We have then, by 3.3, (57), (58) for $n > n_0$

$$(59) \quad I({}_oS_n, \|X\|) = I(T_n, B_n, \|X\|) \leqq (1/m)I(T_n, B_n, f) = (1/m)I({}_oS_n, f).$$

(56) and (59) imply the existence of a finite constant N such that $I({}_oS_n, \|X\|) < N$ for every n. By 3.6 this implies that

$$\liminf I({}_oS_n, f) \geqq I({}_oS_0, f),$$

in contradiction with (56).

3.8. Theorem. *If the admissible integrand f is independent of x, that is, if $f=f(X)$, then the functional $I({}_oS, f)$ is lower semi-continuous in ${}_o\mathfrak{C}$ at every surface ${}_oS_0 \in {}_o\mathfrak{C}$ that satisfies condition* (c) *with respect to f.*

Proof. Checking through the proof of the theorem in 3.5, where the stronger condition (c+) was assumed, we find (cf. 2.5) that the positivity of the E-function was used solely to obtain an estimate for the difference

$$f(x_n(u), \overline{X}_n(u)) - f(x_0(u_0), \overline{X}_n(u)).$$

Since this difference vanishes in the present case, it is clear that condition (c+) can be replaced by the weaker condition (c).

3.9. In what precedes we worked with *oriented* continuous surfaces. Checking back through our discussion, we find that this restriction served the sole purpose of securing the invariance of the integral $I(T, B, f)$ in the quasi-linear case. From the remarks made in 1.10 it is then clear that *the restriction to oriented surfaces becomes unnecessary if the integrand f satisfies the condition*

$$(60) \qquad f(x, X) = f(x, -X).$$

That is, if (60) holds, then in all that precedes we can use the *Fréchet distance* and *continuous surfaces* instead of the *oriented distance* and *oriented continuous surfaces*. By way of illustration, we state presently the theorems corresponding to those in 3.2, 3.5, 3.7, 3.8. We shall use $\mathfrak{C}$ to denote the class of all continuous surfaces of class K_2.

Theorem. *If f is admissible, then the quantity $I(T, B, f)$ has the same value for all representations of class K_2 of the surface $S \in \mathfrak{C}$ if the admissible integrand f satisfies* (60). *This quantity may be therefore denoted by $I(S, f)$.*

Theorem. *If f is an admissible integrand satisfying* (60), *then the functional $I(S, f)$ is lower semi-continuous on $\mathfrak{C}$ at every surface $S \in \mathfrak{C}$ that satisfies either condition* (+c) *or* (c+) *with respect to f. If in addition f is independent of x, that is, if $f=f(X)$, then we have lower semi-continuity at every surface $S \in \mathfrak{C}$ that satisfies condition* (c) *with respect to f.*

3.10. Definition. *If an oriented continuous surface ${}_oS$ admits of a representation (T, B) which is quasi-linear, then ${}_oS$ will be called an oriented polyhedron and will be denoted by ${}_o\mathfrak{P}$. Clearly ${}_o\mathfrak{P} \in {}_o\mathfrak{C}$, since obviously every quasi-linear triple is of class K_2.*

3.11. Definition. *If a (non-oriented) continuous surface S admits of a representation which is quasi-linear, then S will be called a polyhedron and will be denoted by $\mathfrak{P}$. Clearly $\mathfrak{P} \in \mathfrak{C}$, since every quasi-linear triple is of class K_2.*

3.12. Definition. *Let ${}_oS$ be an oriented continuous surface. Its Lebesgue area $L({}_oS)$ is defined as the greatest lower bound of*

$$\liminf I({}_o\mathfrak{P}_n, \|X\|)$$

for all sequences ${}_o\mathfrak{P}_n$ such that ${}_o\mathfrak{P}_n \to {}_oS$.

3.13. Definition. *Let S be a (non-oriented) continuous surface. Its Lebesgue area $L(S)$ is defined as the greatest lower bound of*

$$\liminf I(\mathfrak{P}_n, \|X\|),$$

for all sequences $\mathfrak{P}_n$ such that $\mathfrak{P}_n \to S$.

3.14. Theorem. *Let ${}_oS$ be of class K_1. Then for every representation (T, B) of class K_1 of ${}_oS$ we have*

$$L({}_oS) \geqq I(T, B, \|X\|). \tag{61}$$

The sign of equality holds if and only if (T, B) is of class K_2([24]).

Proof. Let ${}_o\mathfrak{P}_n$ be any sequence such that ${}_o\mathfrak{P}_n \to {}_oS$, and let (T_n, B_n) be a quasi-linear representation of ${}_o\mathfrak{P}_n$. Then

$${}_od[(T_n, B_n), (T, B)] \to 0.$$

Obviously *every* triple of class K_1 satisfies condition (c+) with respect to the integrand $\|X\|$. Hence, by 2.6,

$$\liminf I(T_n, B_n, \|X\|) \geqq I(T, B, \|X\|).$$

Since this holds for every sequence ${}_o\mathfrak{P}_n$ such that ${}_o\mathfrak{P}_n \to {}_oS$, and since

$$I(T_n, B_n, \|X\|) = I({}_o\mathfrak{P}_n, \|X\|)$$

by (3.3), the inequality (61) follows. *Suppose now that*

$$L({}_oS) = I(T, B, \|X\|). \tag{62}$$

It follows from the definition of $L({}_oS)$ that we have a sequence ${}_o\mathfrak{P}_n^*$ such that

$${}_o\mathfrak{P}_n^* \to {}_oS \tag{63}$$

and

$$I({}_o\mathfrak{P}_n^*, \|X\|) \to L({}_oS). \tag{64}$$

Let (T_n^*, B_n^*) be a quasi-linear representation of ${}_o\mathfrak{P}_n^*$. We have then, by (63),

([24]) This theorem could also be inferred from Radó [2].

$$(65)\qquad {}_o d[(T_n^*, B_n^*), (T, B)] \to 0,$$

while (64) and (62) imply that (cf. 3.3)

$$(66)\qquad I(T_n^*, B_n^*, \|X\|) \to I(T, B, \|X\|).$$

By (65), (66) the triple (T, B) is of class K_2 (cf. 1.19).

Suppose conversely that (T, B) *is of class* K_2. We have then, by definition, a sequence of quasi-linear triples $(\overline{T}_n, \overline{B}_n)$ such that

$$(67)\qquad {}_o d[(\overline{T}_n, \overline{B}_n), (T, B)] \to 0,$$

and

$$(68)\qquad I(\overline{T}_n, \overline{B}_n, \|X\|) \to I(T, B, \|X\|).$$

The quasi-linear triple $(\overline{T}_n, \overline{B}_n)$ is a representation of an oriented polyhedron ${}_o\overline{\mathfrak{P}}_n$. Using (3.3), (1.22), the relations (67), (68) can be rewritten in the form

$$(69)\qquad {}_o\overline{\mathfrak{P}}_n \to {}_oS,$$

$$(70)\qquad I({}_o\overline{\mathfrak{P}}_n, \|X\|) \to I(T, B, \|X\|).$$

By the definition of $L({}_oS)$, (69) and (70) imply that

$$(71)\qquad I(T, B, \|X\|) \geqq L({}_oS).$$

(71) and (61) (which we have already proved) imply that

$$L({}_oS) = I(T, B, \|X\|).$$

3.15. Theorem. *The oriented continuous surface ${}_oS$ is of class K_2 if and only if it admits of a representation*

$$T:\ x(u), \qquad u \in B,$$

such that the following conditions hold.

(a) *The Jacobians $X^1(u)$, $X^2(u)$, $X^3(u)$ exist a.e. in B^0 and are summable there.*

(b) $L({}_oS) = \iint_{B_0} \|X(u)\| du$.

Briefly: ${}_oS$ is of class K_2 if and only if it admits of a representation where the Lebesgue area is given by the usual integral formula.

This theorem is merely a rewording of the second half of the theorem in 3.14.

3.16. Theorem. *If (T, B) is a quasi-linear representation of the oriented polyhedron ${}_o\mathfrak{P}$, then*

$$L({}_o\mathfrak{P}) = I(T, B, \|X\|).$$

Briefly: The Lebesgue area of an oriented polyhedron is equal to its area in the elementary sense.

This theorem is merely a very special case of 3.14, since every quasi-linear triple is of class K_2.

3.17. Similar theorems hold for non-oriented surfaces. Since the proofs are entirely analogous to those in 3.14, 3.15, 3.16, we only state the results in the following theorems.

3.18. THEOREM. *Let S be of class K_1. Then for every representation (T, B) of class K_1 of S we have*

$$L(S) \geqq I(T, B, \|X\|).$$

The sign of equality holds if and only if (T, B) is of class K_2.

3.19. THEOREM. *A continuous surface S is of class K_2 if and only if it admits of a representation*

$$T: x(u), \quad u \in B$$

such that the following conditions hold.

(a) *The Jacobians $X^1(u)$, $X^2(u)$, $X^3(u)$ exist a.e. in B^0 and are summable there.*

(b) $L(S) = \iint_{B^0} \|X(u)\| du$.

Briefly: S is of class K_2 if and only if it admits of a representation where the Lebesgue area of S is given by the usual integral formula.

3.20. THEOREM. *If (T, B) is a quasi-linear representation of a (non-oriented) polyhedron $\mathfrak{P}$, then*

$$L(\mathfrak{P}) = I(T, B, \|X\|).$$

Briefly: The Lebesgue area of a polyhedron is equal to its area in the elementary sense.

3.21. *In conclusion, we want to compare our surfaces of class K_2 with the surfaces used by McShane and Morrey in their researches*(25). For brevity, we shall restrict our remarks to *non-oriented* continuous surfaces, and we shall follow the presentation of Morrey.

3.22. According to Morrey, a continuous surface S is of class L if it admits of a representation

$$T_0: \quad x_0(u), \qquad u \in R_0, \tag{72}$$

where R_0 is the square given by

$$R_0: \quad 0 \leqq u^1 \leqq 1, \qquad 0 \leqq u^2 \leqq 1,$$

such that the following conditions are satisfied(26).

(25) See McShane [1, 2] and Morrey [1].

(26) For convenience, we have split the two conditions (i), (ii) of Morrey [1, p. 701], into three conditions (a), (b), (c).

(a) $x_0^i(u)$, $i=1, 2, 3$, is absolutely continuous in the sense of Tonelli.

(b) The Jacobians

$$\frac{\partial(\overset{2}{x_0}, \overset{3}{x_0})}{\partial(u^1, u^2)}, \qquad \frac{\partial(\overset{3}{x_0}, \overset{1}{x_0})}{\partial(u^1, u^2)}, \qquad \frac{\partial(\overset{1}{x_0}, \overset{2}{x_0})}{\partial(u^1, u^2)}$$

are summable in R_0.

(c) For every rectangle R: $a^1 \leqq u^1 \leqq b^1$, $a^2 \leqq u^2 \leqq b^2$, completely interior to R_0, we have

$$\iint_R | X_0^i(u) - X_h^i(u) | \, du \xrightarrow[h \to 0]{} 0, \qquad i = 1, 2, 3,$$

where $X_0^1(u)$, $X_0^2(u)$, $X_0^3(u)$ are the Jacobians corresponding to the given triple T_0, while $X_h^1(u)$, $X_h^2(u)$, $X_h^3(u)$ are the Jacobians corresponding to the triple

$$T_h: \quad x_h(u), \qquad u \in R_h,$$

where

$$\overset{i}{x_h}(u) = \frac{1}{4h^2} \int_{-h}^{h} \int_{-h}^{h} \overset{i}{x_0}(\overset{1}{u} + \overset{1}{v}, \overset{2}{u} + \overset{2}{v}) d\overset{1}{v} d\overset{2}{v}, \qquad i = 1, 2, 3,$$

and R_h is the rectangle

$$R_h: \quad h \leqq u^1 \leqq 1 - h, \qquad h \leqq u^2 \leqq 1 - h.$$

If these conditions are satisfied, then the representation (72) and the triple T_0 will be also said to be of class L.

3.23. *Suppose the representation* (72) *is of class L.* Let us define, for every positive integer n, a rectangle

$$R^{(n)}: \quad \frac{1}{n} \leqq u^1 \leqq 1 - \frac{1}{n}, \qquad \frac{1}{n} \leqq u^2 \leqq 1 - \frac{1}{n}. \tag{73}$$

Clearly then

$$_0d[(T_0, R^{(n)}), (T_0, R_0)] \xrightarrow[n \to \infty]{} 0, \tag{74}$$

and

$$\iint_{R^{(n)}} \|X_0(u)\| du \xrightarrow[n \to \infty]{} \iint_{R_0} \|X_0(u)\| du. \tag{75}$$

For fixed n, the rectangle $R^{(n)}$ will be completely interior to the rectangle R_h if h is sufficiently small. By a well known property of integral means, we have $x_h^i(u) \to_{h \to 0} x_0^i(u)$, $i=1, 2, 3$, uniformly in $R^{(n)}$ for fixed n. Combining this remark with condition (c) in 3.22, we find that for every n we have an $h_n > 0$ such that

$$ {}_0d[(T_{h_n}, R^{(n)}), (T_0, R^{(n)})] < \frac{1}{n}, \tag{76} $$

$$ \iint_{R^{(n)}} \|X_0(u) - X_{h_n}(u)\| du < \frac{1}{n} \cdot \tag{77} $$

Since for fixed n and small h the functions $x_h^i(u)$ are continuous in $R^{(n)}$ together with their partial derivatives of the first order, we have by a familiar reasoning a quasi-linear triple $(\overline{T}^{(n)}, R^{(n)})$ such that

$$ {}_0d[(\overline{T}^{(n)}, R^{(n)}), (T_{h_n}, R^{(n)})] < \frac{1}{n}, \tag{78} $$

$$ \iint_{R^{(n)}} \|X_{h_n}(u) - \overline{X}^{(n)}(u)\| du < \frac{1}{n}, \tag{79} $$

where $\overline{X}^{(n)}(u)$ is the vector whose components are the Jacobians corresponding to $(\overline{T}^{(n)}, R^{(n)})$. Combining (74)–(79), we now obtain the relations

$$ {}_0d[(\overline{T}^{(n)}, R^{(n)}), (T_0, R_0)] \xrightarrow[n\to\infty]{} 0, $$

$$ \iint_{R^{(n)}} \|\overline{X}^{(n)}(u)\| du \xrightarrow[n\to\infty]{} \iint_{R_0} \|X_0(u)\| du. $$

Since the triples $(\overline{T}^{(n)}, R^{(n)})$ are quasi-linear, the last two relations imply that the triple (T_0, R_0) is of class K_2 (cf. 1.19).

3.24. In other words: *Every triple (T_0, R_0) which is of class L in the sense of Morrey is of class K_2 in our sense. In fact, in establishing this result we did not use condition* (a) *in* 3.22 *at all*, as a glance through 3.23 shows. *Hence, every triple (T_0, R_0) which satisfies conditions* (b) *and* (c) *in* 3.22, *is of class K_2 in our sense.*

3.25. *Morrey proved that if conditions* (a), (b), (c) *in* 3.22 *are satisfied, then the Lebesgue area of the surface S determined by the triple (T_0, R_0) is given by the usual integral formula.* Condition (a), requiring that the coordinate functions $x_0^i(u)$ be *absolutely continuous in the Tonelli sense*, played an essential part in his proof, and consideration of the case of surfaces given in non-parametric form seemed to suggest that condition (a) could not be dispensed with. However the present author found some years later that *condition* (a) *could be replaced by the weaker condition requiring only bounded variation in the sense of Tonelli* on the part of $x_0^i(u)$(27). *At present we see that even this weaker condition is unnecessary.* Indeed, by 3.24, conditions (b) and (c) in 3.22 are sufficient to insure that (T_0, R_0) is of class K_2. Hence, by 3.14, the Lebesgue area of the surface represented by (T_0, R_0) is given by the usual integral formula as soon as (T_0, R_0) satisfies only conditions (b) and (c) in 3.22.

(27) See Radó [1].

References

S. Banach

1. *Sur une classe de fonctions d'ensemble*, Fundamenta Mathematicae, vol. 6 (1924), pp. 170–188.

G. A. Bliss

1. *The Weierstrass E-function for problems of the calculus of variations in space*, these Transactions, vol. 15 (1914), pp. 369–378.

P. Franklin and N. Wiener

1. *Analytical approximations to topological transformations*, these Transactions, vol. 28 (1926), pp. 762–785.

E. J. McShane

1. *On the semi-continuity of double integrals in the calculus of variations*, Annals of Mathematics, (2), vol. 33 (1932), pp. 460–484.
2. *Integrals over surfaces in parametric form*, Annals of Mathematics, (2), vol. 34 (1933), pp. 815–838.

C. B. Morrey

1. *A class of representations of manifolds*, I, American Journal of Mathematics, vol. 55 (1933), pp. 683–707.

T. Radó

1. *A remark on the area of surfaces*, American Journal of Mathematics, vol. 58 (1936), pp. 598–606.
2. *On the derivative of the Lebesgue area of continuous surfaces*, Fundamenta Mathematicae, vol. 30 (1938), pp. 34–39.
3. *On absolutely continuous transformations in the plane*, Duke Mathematical Journal, vol. 4 (1938), pp. 190–221.
4. *On a lemma of McShane*, Annals of Mathematics, (2), vol. 42 (1941), pp. 73–83.

The Ohio State University,
Columbus, Ohio.

MANIFOLDS WITHOUT CONJUGATE POINTS

BY

MARSTON MORSE AND GUSTAV A. HEDLUND

1. **Introduction.** If a closed two-dimensional Riemannian manifold M is homeomorphic to a sphere or to the projective plane and A is any point of M, there exists a geodesic passing through A with a point on it conjugate to A (cf., e.g., Myers [1, p. 48, Corollary 2])(1). There exists a closed two-dimensional Riemannian manifold of any other given topological type such that no geodesic on the manifold has on it two mutually conjugate points. The simplest examples of these are manifolds of vanishing Gaussian curvature in the case of the torus and the Klein bottle, and manifolds of constant negative curvature in the remaining cases. In all these particular examples the differential equations defining the geodesics can be integrated and the properties in the large of the geodesics can be determined. In the case of the flat torus or flat Klein bottle the geodesics are either periodic or recurrent but not periodic. In the case of closed manifolds of constant negative curvature the behavior of the geodesics is much more complex, but among other types there exist transitive geodesics.

In this paper our starting point will be the assumption that we have a closed two-dimensional Riemannian manifold M such that no geodesic on M has on it two mutually conjugate points. Our aim is to determine what properties of M or of the geodesics on M must follow from this hypothesis. The possibility that M be homeomorphic to the sphere or to the projective plane is thereby eliminated and it is necessary to divide the possible closed manifolds into two classes. The first class is made up of manifolds homeomorphic to the torus or to the Klein bottle; the second class is made up of the remaining possible manifolds.

In Part I we study manifolds of the first class. Here the universal covering surface of M is a plane Θ provided with a metric which satisfies certain group properties. It is shown that the hypothesis that there are no two mutually conjugate points on any geodesic implies that the geodesics in Θ behave in numerous respects like straight lines. Each unending geodesic g is the topological image of a straight line. There exists a constant R determined by M such that any unending geodesic lies between two parallel straight lines at a distance apart not exceeding R. Two geodesics can intersect in at most one point. A parallelism property holds in that, if g is any unending geodesic and P is any point not on g, there exists exactly one geodesic passing through P and not intersecting g.

Presented to the Society, September 5, 1941; received by the editors February 11, 1941.

(1) References will be found in the bibliography at the end of the paper.

There is the possibility that the hypothesis concerning the nonexistence of conjugate points implies that the Gaussian curvature vanishes identically. The authors have not been able to verify or disprove this conjecture.

However, it will be shown that an additional hypothesis concerning the nonexistence of focal points is sufficient to insure the identical vanishing of the Gaussian curvature.

This additional hypothesis as well as that concerning the nonexistence of conjugate points is implied by the hypothesis that, if g is any geodesic in Θ and P is any point not on g, there is just one geodesic passing through P and orthogonal to g.

In Part II we consider manifolds of the second class, as well as certain manifolds which are not closed. Here the universal covering surface is the interior of the unit circle. The hypothesis that no geodesic have on it two mutually conjugate points is fulfilled by a large class of manifolds and, in particular, by those with everywhere negative curvature.

The question with which we shall be principally concerned is that of the existence of transitive geodesics. The existence of such geodesics on closed orientable manifolds of genus greater than one has been proved under various hypotheses. The first proof which did not assume that the Gaussian curvature of the manifold was everywhere negative was given by Morse (cf. Morse [2]), who showed the existence of transitive geodesics under the hypothesis of *uniform instability*. Subsequently, Hedlund (cf. Hedlund [2]) proved that such geodesics exist on an extended class of manifolds, provided a different condition, which we term *ray instability*, is fulfilled. Both uniform instability and ray instability imply that no geodesic has on it two mutually conjugate points.

In the present paper we show that for a large class of two-dimensional manifolds the hypothesis that no geodesic has on it two mutually conjugate points is sufficient to imply the existence of transitive geodesics.

The attainment of the new results of both parts of the paper is made possible by use of the well known recurrence theorem of Poincaré which states that in a dynamical system with a bounded invariant integral almost all the motions are stable in the sense of Poisson.

Part I. Manifolds of euclidean type

2. A class of simply-connected Riemannian manifolds of euclidean type. We consider the quadratic form

$$ds^2 = F^2(x, y)(dx^2 + dy^2) \tag{2.1}$$

where $F(x, y)$ is of class C^3 and

$$0 < a \leqq F(x, y) \leqq b, \tag{2.2}$$

in the (x, y)-plane Θ. This two-dimensional Riemannian manifold will be denoted by $M(F)$ and termed a *manifold of euclidean type*.

The length of a rectifiable curve γ on $M(F)$ will be its length as determined by the metric (2.1) and will be denoted by $L(\gamma)$. The geodesics corresponding to (2.1) are of class C^2 in terms of the arc length as parameter and in terms of suitably chosen parameters giving the initial conditions.

The manifold $M(F)$ is complete in the sense of Hopf and Rinow (cf. Hopf and Rinow [1]) and corresponding to a given pair of points P and Q of Θ there exists a geodesic segment joining P and Q which affords an absolute minimum of length relative to all rectifiable curves joining P and Q. Following Morse (cf. Morse [1]) we term such a geodesic segment a *geodesic segment of class A*. We term the length of a class A geodesic segment joining P and Q the *distance* between P and Q and denote it by $D(P, Q)$. It is easily shown that the metric $D(P, Q)$ satisfies the usual conditions (cf. Hausdorff [1, p. 94]) which are imposed on a metric.

A geodesic segment of class A is necessarily a simple curve. For, if the point P were a multiple point of the segment, we would have a geodesic segment of class A with identical end points and such a segment obviously could not be of class A.

A *geodesic ray r* is a geodesic with an initial point P which in terms of its arc length s measured from P is the continuous image of the half-line, $0 \leqq s < \infty$. A geodesic ray is of *class A*, if every finite segment of it is a class A geodesic segment. It is evident that a geodesic ray of class A can have no multiple points and is the topological image of a half-line. As to the existence of class A geodesic rays, it is known (cf. Rinow [1]) that, if P is any point of Θ, there exists a class A geodesic ray with initial point P.

An *unending geodesic* is a geodesic which in terms of its arc length s is the continuous image of the whole s-axis, $-\infty < s < \infty$. An unending geodesic g is said to be of *class A* if each finite segment of g is of class A. A class A unending geodesic can have no multiple points and is the topological image of a straight line. Under the conditions thus far imposed on $M(F)$ we cannot infer the existence of unending geodesics of class A on $M(F)$. As a matter of fact, it is not difficult to construct manifolds $M(F)$ on which there are no unending geodesics of class A.

If we denote the euclidean distance (i.e., the distance when $F(x, y) \equiv 1$) between P and Q by $E(P, Q)$, the following lemma can be stated.

LEMMA 2.1. *If P and Q are arbitrary points of* Θ,

$$aE(P, Q) \leqq D(P, Q) \leqq bE(P, Q).$$

For let γ be a class A geodesic segment joining P and Q. With the aid of (2.2) we find that

$$D(P, Q) = \int_\gamma F(\dot{x} + \dot{y})^{1/2} dt \geqq a \int_\gamma (\dot{x}^2 + \dot{y}^2)^{1/2} dt \geqq aE(P, Q),$$

and the first inequality of the lemma is proved. If we let λ be the E-line (euclidean line) segment joining P and Q, we have

$$D(P, Q) = \int_\gamma ds \leqq \int_\lambda ds \leqq b \int_\lambda (\dot{x}^2 + \dot{y}^2)^{1/2} dt = bE(P, Q),$$

and the proof of the lemma is complete.

Let r be a geodesic ray, s the arc length on r measured from the initial point, $P(s)$ the point of r determined by s, and 0 the origin. The geodesic ray r will be said to *recede to infinity* if $\lim_{s\to+\infty} E(0, P(s)) = +\infty$. An unending geodesic g will be said to *recede to infinity* if each of the geodesic rays into which g is divided by any point recedes to infinity.

As a rather evident consequence of Lemma 2.1, we can state the following theorem.

THEOREM 2.1. *Every class A geodesic ray and every class A unending geodesic on $M(F)$ recedes to infinity.*

We shall subsequently be concerned with the question of the extent to which geodesic segments, geodesic rays, or unending geodesics of class A behave like euclidean line segments, half-lines or whole lines, respectively. To that end we introduce the following well known definitions.

If U is a set of points in Θ and P is a point of Θ, we define the *distance* of the point P from the set U, or $D(P, U)$, by

$$D(P, U) = \underset{u\in U}{\text{g.l.b.}}\ D(P, u).$$

If U and V are point sets in Θ we define (cf. Hausdorff [1, p. 146]) the *type-distance* between these sets, or $D(U, V)$, by

$$D(U, V) = \max\left(\underset{u\in U}{\text{l.u.b.}}\ D(u, V);\ \underset{v\in V}{\text{l.u.b.}}\ D(v, U)\right).$$

If either of the bounds on the right does not exist, $D(U, V) = +\infty$.

Two sets U and V will be said to be *of the same type*, or each will be said to be *of the type of* the other, if $D(U, V)$ is finite.

If there were a uniform upper bound R of the type-distance between a class A geodesic segment and the E-line (euclidean line) segment with the same end points, we could infer the existence of a large class of unending geodesics of class A. As has been stated, the hypotheses made up to this point do not imply the existence of any class A unending geodesics, so that additional restrictions must be imposed to secure the existence of the constant R

3. **Doubly-periodic Riemannian manifolds.** Let T_1 and T_2 be translations of Θ into Θ such that the fixed lines of T_1 and T_2 are not identical. Let G be the doubly-periodic group with T_1 and T_2 as generators.

The condition that $F(x, y)$ be invariant under the translations of G will be

denoted by (I). *A two-dimensional Riemannian manifold* $M(F)$ *which satisfies* (I) *will be termed a doubly-periodic Riemannian manifold and denoted by* $M(F, \mathrm{I})$.

The following theorem has been proved by Hedlund. (Cf. Hedlund [1, p. 731]. The groups of translations considered in Hedlund [1] were made up only of translations parallel to the axes, but the arguments used in the proofs of the theorems require no essential modifications in order to apply to the group G of the present paper. The corresponding theorem of Hedlund [1] is stated in terms of E-distance, but Lemma 2.1 permits restatement in terms of the type-distance defined here.)

THEOREM 3.1. *Corresponding to a given* $M(F, \mathrm{I})$ *there exists a finite constant* R *such that the type-distance between any class* A *geodesic segment on* $M(F, \mathrm{I})$ *and the* E*-line segment with the same end points cannot exceed* R.

As a consequence of this theorem, the following can be deduced (cf. Hedlund [1, p. 732]).

THEOREM 3.2. *Each unending geodesic of class* A *on an* $M(F, \mathrm{I})$ *is of the type of some* E*-line. Conversely, corresponding to an arbitrary* E*-line* L, *there is at least one unending class* A *geodesic of the type of* L.

We will term a closed half-line of Θ a euclidean ray or E-ray.

THEOREM 3.3. *Each geodesic ray* r *of class* A *on an* $M(F, \mathrm{I})$ *is of the type of a unique* E*-ray* l *with the same initial point and the type-distance between* l *and* r *does not exceed the constant* R *of Theorem* 3.1. *Conversely, corresponding to any* E*-ray* l *there exists a class* A *geodesic ray* r *on* $M(F, \mathrm{I})$ *of the same type and with the same initial point as* l. *The type distance between* l *and* r *does not exceed the constant* R.

Let r be a class A geodesic ray with initial point P_0 and let s be the arc length on r measured from P_0. Let P_n be the point on r such that the value of s determining P_n is n. According to Theorem 2.1, r recedes to infinity and consequently as n becomes infinite the E-distance P_0P_n must become infinite. If the sequence $0<n_1<n_2< \cdots$ is properly chosen, the sequence of E-rays $l_1, l_2, \cdots$, where l_i has the initial point P_0 and passes through P_{n_i}, will have a unique limiting E-ray l with initial point P_0. According to Theorem 3.1, the type-distance between the segment $P_0P_{n_i}$ of r and the E-line segment $P_0P_{n_i}$ does not exceed R. It follows that the type-distance between l and r does not exceed R. Thus the existence of at least one E-ray l corresponding to r and with the stated properties has been proved.

To show that l is unique, suppose a second E-ray l' exists with initial point P_0 and such that r is also of the type of l'. Since l and l' are both of the type of r, it would follow that l and l' are of the same type and this is evidently not the case.

Conversely, given an E-ray l, let $P_0, P_1, \cdots$ be an ordered sequence of points on l such that P_0 is the initial point of l and the E-distance P_0P_n becomes infinite with n. If g_n is a class A geodesic segment joining P_0 and P_n, it follows from Theorem 3.1 that g_n cannot recede a distance exceeding R from the E-line segment P_0P_n. If g_n is oriented so that P_0 is its initial point, let e_n denote the element of g_n at P_0. That is, e_n is a triple of numbers (x_0, y_0, ϕ_n), where (x_0, y_0) are the coordinates of P_0 and where ϕ_n, with $0 \leqq \phi_n \leqq 2\pi$, is the angular coordinate at P_0 determined by the direction of g_n at P_0. The sequence $\phi_1, \phi_2, \cdots$ contains a subsequence converging to some value ϕ and we say that the corresponding subsequence of $e_1, e_2, \cdots$ converges to the element $e(x_0, y_0, \phi)$. The geodesic ray r with initial element e must be of class A, for it is the limit of class A geodesic segments g_n. Since no point of g_n can be at a distance exceeding R from the E-line segment P_0P_n, it follows that no point of r can be at a distance from l exceeding R.

On the other hand, if Q is any point of l, Q is a point of P_0P_n for n greater than some properly chosen integer N. It follows from Theorem 3.1 that for $n > N$, g_n has on it a point Q_n such that $D(Q, Q_n) \leqq R$. Since g_n is of class A, the length of the segment P_0Q_n of g_n is less than a constant independent of n. In fact

$$D(P_0, Q_n) \leqq D(P_0, Q) + D(Q, Q_n) \leqq D(P_0, Q) + R.$$

It follows that the sequence $Q_1, Q_2, \cdots$ must contain a subsequence which converges to a point Q' of r. But then $D(Q, Q') \leqq R$, and no point of l can be at a distance exceeding R from r. Thus r and l are of the same type with type-distance not exceeding R and the proof of Theorem 3.3 is complete.

THEOREM 3.4. *Let g_1 and g_2 be class A unending geodesics on an $M(F, \mathrm{I})$ such that g_1 and g_2 are not identical and intersect in a point P. Corresponding to $\epsilon > 0$, there exists a $\delta > 0$ such that, if g_1' and g_2' are those point sets of g_1 and g_2, respectively, which are at distance at least ϵ from P, the distance between any point of g_1' and any point of g_2' cannot be less than δ.*

The proof of this theorem is essentially the same as that of Theorem 6 of Morse [1]. If Theorem 3.4 were not true, there would exist an $\epsilon > 0$, a sequence $\delta_1 > \delta_2 > \cdots$, with $\lim_{i\to+\infty} \delta_i = 0$, and a sequence of point pairs (M_1, N_1), $(M_2, N_2), \cdots$, M_i on g_1, N_i on g_2, with both M_i and N_i at distance at least ϵ from P, and such that $\lim_{i\to+\infty} D(M_i, N_i) = 0$. By proper choice of subsequences we can assume that P does not separate on g_1 any two points M_i and M_j, nor any two points N_i and N_j on g_2. Let Q be a point on g_1, distinct from P and such that on g_1, P separates Q and any point M_i. Then the length of the broken geodesic $\bar{g}_i$ consisting of the segment QP of g_1 and the segment PN_i of g_2 does not exceed the length of the geodesic segment QM_i of g_1 by more than δ_i. But the broken geodesic $\bar{g}_i$ has a corner at P and the points Q and N_i can be joined by a geodesic segment the length of which is less than

the length of $\bar{g}_i$ by a fixed amount which is independent of i. Since $\lim_{i\to+\infty} \delta_i = 0$, we infer that for i sufficiently large the segment QM_i of g_1 cannot be the shortest geodesic segment joining Q and M_i, contrary to the hypothesis that g_1 is of class A.

The proof of the theorem is complete.

A set of points U of the plane Θ will be said to be periodic with period (m, n), if U is invariant under the translation $T_1^m T_2^n$, $m^2+n^2 \neq 0$, of G.

The period (m, n), $m>0$, is a *primitive period* if m and n are relatively prime. The period $(0, n)$ is a *primitive period* if $n=1$. Corresponding to any period (m, n), there exists a unique primitive period (μ, ν) such that $m=k\mu$, $n=k\nu$, k integral. We term (μ, ν) the *primitive period corresponding to* (m, n). It is easily shown that if an E-line has the period (m, n), it has all the periods and only the periods $(i\mu, i\nu)$, $i=\pm 1, \pm 2, \cdots$, where (μ, ν) is the primitive period corresponding to (m, n).

THEOREM 3.5. *If g is an unending class A periodic geodesic on an $M(F, \mathrm{I})$ and g is of the type of the periodic E-line L with primitive period (μ, ν), then g has all the periods and only the periods $(j\mu, j\nu)$, $j=\pm 1, \pm 2, \cdots$.*

For suppose that g has the period (m, n). Then g is invariant under the transformation $T_1^m T_2^n$. We show that $T_1^m T_2^n = T$ transforms L into itself.

Since L and g are of the same type, their type-distance is less than a fixed constant C. The metric $D(P, Q)$ is invariant under T and all its powers, so that the type-distance between $T^k(L)$ and $T^k(g)$, k any positive integer, is less than C. But $T^k(g)=g$, and if $T(L)$ were not L, the type-distance between $T^k(L)$ and L would increase without limit as k becomes infinite. It would follow that for k sufficiently large, the type-distance between $T^k(g)=g$ and $T^k(L)$ would exceed C. But this type-distance is less than C and we infer that $T(L)=L$. Since L has only periods of the form $(j\mu, j\nu)$, j integral, it follows that $m=i\mu$, $n=i\nu$, i a nonvanishing integer.

It remains to show that g actually has (μ, ν) among its periods. If we let $T_1^\mu T_2^\nu = \overline{T}$, it follows that $\overline{T}^i = T$. Suppose that g is not invariant under $\overline{T}$. Then $\overline{T}(g)=g'$ is periodic and not identical with g. Moreover

$$T(g') = T(\overline{T}(g)) = \overline{T}(T(g)) = \overline{T}(g) = g'.$$

Thus g and g' are both invariant under T. Hence g and g', both of which are of class A, can have no points in common. The unending geodesic g divides the plane Θ, excluding g, into two open sets Θ_1 and Θ_2 in one of which g' lies. We can assume that g' lies in Θ_2. The geodesic g' divides the plane Θ into two open sets Θ_1' and Θ_2', of which one, and we assume it to be Θ_1', contains g, and hence Θ_1. Then $\overline{T}(\Theta_1)=\Theta_1'$ and $\overline{T}(\Theta_1)$ contains Θ_1 and additional points. But $\overline{T}^2(\Theta_1)=\overline{T}(\Theta_1')$ contains Θ_1 and additional points. In general, $\overline{T}^k(\Theta_1)$ contains Θ_1 and additional points, if $k>0$. Since $\overline{T}^i(g)=g$, it follows that $\overline{T}^i(\Theta_1)=\Theta_1$ and we have a contradiction, if i is positive. Similar arguments apply if i is negative.

The proof of the theorem is complete.

4. **A class of closed orientable manifolds of genus one.** Let $M(F, \mathrm{I})$ be a doubly-periodic Riemannian manifold. Then $F(x, y)$ is invariant under a group G of translations. We note that the assumption that F satisfies (2.2) is now implied by the assumption $F>0$. For F is continuous and assumes all its values in a parallelogram. It follows that F has a positive lower bound and a positive upper bound.

Since F is invariant under the group G, the metric (2.1) is invariant under the transformations of G. If points of the (x, y)-plane which are congruent under a transformation of G are considered identical, there is defined a closed orientable two-dimensional Riemannian manifold $M(F, G)$ of genus one. A geodesic on $M(F, G)$ is *represented* in Θ by an infinite set of congruent geodesics of $M(F)$.

If (x, y) is a point of Θ and $0 \leqq \phi < 2\pi$, the triple of numbers (x, y, ϕ) determines a direction ϕ at the point (x, y) of Θ if we assume that ϕ is measured in the positive sense from a direction parallel to the positive x-axis. We term the point $P(x, y)$ the *point bearing the element* (x, y, ϕ). Let $\mathcal{E}$ denote the set of elements (x, y, ϕ), (x, y) in Θ, $0 \leqq \phi < 2\pi$. We topologize the space $\mathcal{E}$ by considering it as the product of the plane Θ and the unit circle.

A transformation of G transforms a point of $\mathcal{E}$ into a *congruent point* or *element* and we denote by Ω the space obtained by identifying congruent points (elements) of $\mathcal{E}$. The space Ω is the space of elements on $M(F, G)$ and is a three-dimensional torus. A point in Ω is *represented* in Θ by an infinite set of congruent elements.

The directed geodesics on $M(F, G)$ define a continuous flow T_s in the space Ω, where s is the arc length along the geodesics (cf., e.g., Hedlund [3]). If we define measure in Ω by means of the integral

$$\iiint F^2(x, y)\,dx\,dy\,d\phi,$$

the transformation T_s is measure-preserving for each real s. It is evident that $m\Omega < +\infty$.

A *motion* in Ω is any set $T_s p$, with $-\infty < s < +\infty$, and p a point of Ω. A motion in Ω is the set of elements on a directed geodesic on $M(F, G)$ and is represented in Θ by an infinite set of congruent directed unending geodesics of $M(F)$.

A motion in Ω defined by the set $T_s p$ is said to be *stable in the sense of Poisson*, if there exists an infinite sequence of values $\cdots < s_{-1} < s_0 < s_1 < s_2 < \cdots$, with $\lim_{i\to\infty} |s_i| = +\infty$, such that $\lim_{|i|\to+\infty} T_{s_i}(p) = p$. Let M be a motion which is stable in the sense of Poisson and let g be a directed unending geodesic of Θ representing M. If s denotes the directed arc length on g measured from some fixed point P_0, there must exist an infinite sequence $\cdots < s_{-1} < s_0 < s_1 < \cdots$, with $\lim_{|i|\to+\infty} |s_i| = +\infty$, and a sequence of points

$\cdots, Q_{-1}, Q_0, Q_1, \cdots$, each congruent to P_0, such that if P_i denotes the point of g determined by s_i, $\lim_{|n|\to+\infty} D(P_n, Q_n)=0$.

Since $m\Omega<+\infty$, it follows from the well known recurrence theorem of Poincaré (cf. Poincaré [1], Carathéodory [1], E. Hopf [1]) that almost all points of Ω are on motions which are stable in the sense of Poisson. We will make extensive use of this theorem.

5. **Properties of the geodesics on doubly-periodic manifolds which satisfy the non-conjugacy hypothesis.** We make the following definition.

(II) NON-CONJUGACY HYPOTHESIS. *There is no pair of mutually conjugate points on any geodesic on $M(F)$.*

THEOREM 5.1. *A necessary and sufficient condition that $M(F)$ fulfill the non-conjugacy hypothesis is that there be only one geodesic segment joining two given points of $M(F)$.*

If there are no two mutually conjugate points on any geodesic g on $M(F)$, the geodesic rays issuing from a point P of Θ, form a field in Θ in the sense that, if Q is any point of Θ other than P, there is one and only one geodesic ray with initial point P and passing through Q. It follows that there is just one geodesic segment joining two points of $M(F)$.

Conversely, if a geodesic g on $M(F)$ has on it two mutually conjugate points, a segment σ of g containing these mutually conjugate points as interior points is not of class A and the end points of σ can be joined by a class A geodesic segment which differs from σ. Thus, if the geodesic segment joining any two given points of $M(F)$ is unique, the non-conjugacy hypothesis is fulfilled.

THEOREM 5.2. *A necessary and sufficient condition that $M(F)$ fulfill the non-conjugacy hypothesis is that all geodesics on $M(F)$ be of class A.*

The necessity follows at once from Theorem 5.1.

The sufficiency follows as in the proof of Theorem 5.1.

We will denote a manifold $M(F, \mathrm{I})$ which fulfills the non-conjugacy hypothesis by $M(F, \mathrm{I}, \mathrm{II})$.

It follows from Theorems 3.2 and 5.2 that each unending geodesic on an $M(F, \mathrm{I}, \mathrm{II})$ is of the type of an E-line. It is evident that two unending geodesics on $M(F, \mathrm{I}, \mathrm{II})$ which are of the types of non-parallel E-lines must cross and thus they intersect in just one point. The question arises as to whether two unending geodesics on $M(F, \mathrm{I}, \mathrm{II})$ which are of the types of parallel E-lines, or, what is equivalent, of the same type, can cross. As we shall see, the answer is in the negative.

LEMMA 5.1. *Two unending geodesics on an $M(F, \mathrm{I}, \mathrm{II})$ which are of the type of the same periodic E-line cannot intersect.*

For suppose that g_1 and g_2 are intersecting unending geodesics on $M(F, \mathrm{I}, \mathrm{II})$ and each is of the type of the periodic E-line L. Let (μ, ν) be the primitive period of L. It follows from Theorem 5.2 that both g_1 and g_2 are of class A. We consider the various possibilities.

Case I. *g_1 and g_2 are both periodic.* It follows from Theorem 3.5 that both g_1 and g_2 have the period (μ, ν). Since g_1 and g_2 intersect in a point P, it would follow that they intersect in the point $T_1^\mu T_2^\nu(P) \neq P$. Since g_1 aad g_2 are both of class A, they cannot intersect in two points and we infer that the hypothesis of Case I is impossible.

Case II. *One of the pair g_1, g_2 is periodic, but not both are periodic.* It is known (cf. Hedlund [1, Theorem XIII]) that a class A unending geodesic g which is of the type of a periodic E-line L is either periodic or is asymptotic in both its senses to periodic class A geodesics which are of the type of L. It is also known (cf. Hedlund [1, Theorem XV]) that a geodesic g of class A which is asymptotic to a periodic geodesic b of class A cannot cross any class A periodic geodesic of the type of b. These two statements are incompatible with the hypothesis of Case II and we infer that Case II is not possible.

Case III. *Neither g_1 nor g_2 is periodic.* The point P in which g_1 and g_2 intersect divides g_1 (g_2) into two geodesic rays r_1 and s_1 (r_2 and s_2), all four geodesic rays thus determined having the point P as initial point and no two of the four intersecting in any point other than P. Since g_1 and g_2 are of the same type, the four geodesic rays can be divided into two pairs such that members of the same pair are of the same type. We assume that the notation has been so chosen that r_1 and r_2 are of the same type.

The rays r_1 and r_2 divide the points of the plane Θ which are not on these rays into two open connected sets σ and λ which are separated by the set consisting of the points of r_1 and r_2. (This is easily seen by projecting the plane stereographically onto the unit sphere. Then the two rays r_1 and r_2, together with the north pole of the sphere, form a Jordan curve on the sphere.) It follows from Theorem 3.3 that a half-strip Γ (i.e., the set of points between two parallel lines and on one side of a line perpendicular to these two) can be so drawn that r_1 and r_2 lie in Γ. But then one of the sets σ or λ (and we assume that it is σ) lies entirely in Γ, and hence is of the type of r_1 (and r_2).

Let g be a geodesic which passes through P and enters the set σ. Then P divides g into two rays of which one, which we denote by r, lies entirely in σ. It follows from Theorem XIII of Hedlund [1], that either g is periodic and of the type of L or r is asymptotic to a periodic (class A) geodesic b which is of the type of L. The first possibility is ruled out by applying Case II to the pair g and g_1. If the second possibility holds, it follows from Theorem 3.4 that b has on it points in σ. But b must leave σ and hence cross either r_1 or r_2. Again we have Case II by use of g_1 and b if r_1 and b cross, or by use of g_2 and b if r_2 and b cross.

The proof of the theorem is complete.

THEOREM 5.3. *Any unending geodesic g on an $M(F$, I, II) which is of the type of a periodic E-line is periodic.*

For let g be an unending geodesic on $M(F$, I, II) and of the type of the periodic E-line L. Then g is of class A, and if g is not periodic, we have seen (Hedlund [1, Theorem XIII]) that g is asymptotic in both its senses to periodic class A geodesics b and c of the type of L. But a geodesic of class A cannot be asymptotic in both its senses to the same periodic geodesic (cf. Hedlund [1, Theorem XVII]) so that b and c are distinct. Since b and c are of class A and periodic, they can have no points in common. There can be no class A periodic geodesic lying between b and c, for such a geodesic would cross g, which is impossible. But then there must exist two different class A geodesics which lie between b and c and which must intersect (cf. Hedlund [1, Theorem XVI]). In view of Lemma 5.1 this cannot be the case. We infer that g must be periodic and the proof of the theorem is complete.

THEOREM 5.4. *An unending geodesic on an $M(F$, I, II) which passes through two congruent points is periodic.*

For let P and Q be a pair of congruent points of Θ and let g be an unending geodesic passing through P and Q. Let l be the E-ray with initial point P and passing through Q. Then l is part of a periodic E-line L with primitive period (μ, ν). According to Theorem 3.3 there exists a class A geodesic ray r of the type of l and with initial point P. Since all geodesics on $M(F$, I, II) are of class A, the unending geodesic g^*, of which r is a part, is of the type of an E-line and must be of the type of L. From Theorem 5.3 we infer that g^* is periodic and from Theorem 3.5 that g^* has the period (μ, ν). But then g^* must pass through Q and thus is identical with g. It follows that g is periodic.

THEOREM 5.5. *Two unending geodesics on an $M(F$, I, II) which are of the same type cannot intersect.*

For suppose g_1 and g_2 are unending geodesics on $M(F$, I, II) which are of the same type and intersect in the point P. It follows from Theorem 5.1 that g_1 and g_2 can intersect in no other point than P. As in the proof of Lemma 5.1, the point P divides g_1 (g_2) into two geodesic rays r_1 and s_1 (r_2 and s_2), all four geodesic rays thus determined having the point P as initial point and no two intersecting in any point other than P. We can assume the notation has been so chosen that r_1 and r_2 (s_1 and s_2) are of the same type and neither r_1 nor r_2 is of the type s_1 or s_2.

The rays r_1 and r_2 divide the points of Θ which are not on these rays into two open connected sets of which one, which we denote by σ, is of the type of r_1 (or r_2). Similarly the rays s_1 and s_2 divide the plane Θ into two connected open sets of which one, which we denote by λ, is of the type of s_1 (or s_2). The set σ can contain no points of s_1 (or s_2), for if this were the case, σ would contain all points of s_1 (or s_2) except P, and could not be of the type of r_1 (or r_2).

Similarly, λ contains no points of r_1 or r_2. It follows that either λ lies entirely in σ or λ contains no point of σ. If the first were true, λ would be of the type of σ, and thus of the type of r_1 (or r_2). Since this is not the case, we infer that λ contains no point of σ. Similarly σ contains no point of λ.

Any geodesic segment g^* which joins a point T in σ and a point U in λ must cross one of the rays s_1 or s_2, and one of the rays r_1 or r_2. Since g^* can cross g_1, which is made up of r_1 and s_1, in at most one point and, similarly, g_2 in at most one point, g^* must cross both g_1 and g_2. If g is the unending geodesic of which g^* is a segment, since g is of class A, g cannot leave the set $\sigma+\lambda$ except for a finite segment of g^* and consequently g must be of the type of g_1 (or g_2).

Now consider the set $\mathcal{G}$ of oriented geodesics each of which has a finite segment with the initial point in λ and the terminal point in σ. All oriented geodesics on the closed manifold $M(F, G)$ which are represented by geodesics in the set $\mathcal{G}$ form a set $\overline{\mathcal{G}}$ of geodesics on $M(F, G)$. The set of elements on the geodesics of $\overline{\mathcal{G}}$ is evidently an open set α in Ω and since α is an open set it is measurable and of positive measure. According to the Poincaré recurrence theorem the set $\overline{\mathcal{G}}$ must contain a motion (geodesic) $\bar{g}$ which is stable in the sense of Poisson. Let g be an unending geodesic in the set $\mathcal{G}$ and representing $\bar{g}$. If s denotes the directed arc length on g measured from some point P_0, there must exist an infinite sequence $\cdots<s_{-1}<s_0<s_1<s_2<\cdots$ with $\lim_{|n|\to\infty}|s_n|=+\infty$, and a sequence of points $\cdots, Q_{-1}, Q_0, Q_1, \cdots$, each congruent to P_0, and such that, if P_i denotes the point of g determined by s_i, $\lim_{|n|\to+\infty} D(P_n, Q_n)=0$. By application of Theorem 3.4 to each of the pairs g, g_1 and g, g_2, we infer that infinitely many of the points $\cdots, Q_{-1}, Q_0, Q_1, \cdots$ are in each of the sets σ and λ. In particular, suppose that Q_μ is in λ and Q_ν is in σ. Then according to Theorem 5.4 the geodesic g_P determined by the congruent points Q_μ and Q_ν is periodic. The periodic geodesic g_P is of the type of g_1 (or g_2), and it follows that g_1 and g_2 are of the type of the same periodic E-line L. But according to Lemma 5.1, it is impossible for g_1 and g_2 to intersect, as we have assumed.

The proof of Theorem 5.5 is complete.

THEOREM 5.6. *If g is an unending geodesic on an $M(F, \mathrm{I}, \mathrm{II})$ and P is a point not on g, there exists one and only one unending geodesic on $M(F, \mathrm{I}, \mathrm{II})$ which passes through P and does not intersect g.*

It follows from Theorem 5.2 that g is of class A and thus according to Theorem 3.2, g is of the type of an E-line L. If l denotes an E-ray with initial point P and of the type of one of the rays into which L is divided by a point Q, it follows from Theorem 3.3 that there exists a geodesic ray r with initial point P and of the type of l. If g^* is the unending geodesic of which r is a subset, g^* is of class A, of the type of an E-line, and hence of the type of L. Since P is not on g, g^* is not identical with g, and since both g and g^* are of the type of L, they are of the same type. We infer from Theorem 5.5 that g and g^* cannot intersect.

If there were two unending geodesics passing through P and each of the type of g, we would have two intersecting geodesics of the same type, contradictory to Theorem 5.5.

The proof of the theorem is complete.

The preceding theorem can be stated as follows.

THEOREM 5.7. *The geodesics on an $M(F, \mathrm{I}, \mathrm{II})$ can be grouped into $\aleph$ families, the members of any one family forming a field in Θ, and any two members from different families intersecting in just one point.*

For let L be an E-line. There exists an unending geodesic g of the type of L. If P is any other point of Θ not on g, it follows from Theorem 5.6 that there is just one unending geodesic g_P passing through P and not intersecting g. The geodesic g_P is of the type of g and hence of the type of L. Thus the set G_L of unending geodesics of the type of L cover the plane Θ. But according to Theorem 5.5, no two members of the set G_1 can intersect and thus the set G_L forms a field in Θ. No two of the E-lines passing through the origin are of the same type, and the existence of $\aleph$ families is evident. The last statement of the theorem follows from the fact that unending geodesics which are not of the same type must intersect, and since all geodesics are of class A there can be at most one point of intersection.

6. **Non-conjugacy and closed orientable surfaces of genus one.** Let M be a two-dimensional Riemannian manifold of class C^3 of the topological type of a torus. That is, M is a two-dimensional topological manifold which is homeomorphic to a torus; with each point p of M there is associated a neighborhood N_p which is mapped by a homeomorphism into the interior of the unit circle, $x^2+y^2<1$, such that the transformation T defined by overlapping neighborhoods is of class C^4 with nonvanishing Jacobian; a metric is assigned to each neighborhood N_p by assigning a quadratic differential form $ds^2=f_p(x, y)(dx^2+dy^2)$ to the unit circle, $f_p(x, y)$ being of class C^3 and positive, and such that the transformation T defined by overlapping neighborhoods transforms one of the corresponding quadratic forms into the other.

The manifold M is a Riemann surface (cf. Koebe [1]) and its universal covering surface can be mapped conformally into the plane Θ, thereby defining a metric of the form (2.1) which satisfies (2.2) and (I). If we denote this latter manifold by $M(F, \mathrm{I})$, M is obtained from $M(F, \mathrm{I})$ by identification of congruent points. Thus M is the manifold $M(F, G)$.

If there are no two mutually conjugate points on any geodesic on M, the corresponding Riemannian manifold defined in Θ is an $M(F, \mathrm{I}, \mathrm{II})$. It follows that the theorems of §5 are applicable to the surface M. We state the principal results in the following theorem.

THEOREM 6.1. *Let M be an orientable two-dimensional Riemannian manifold of genus one and of class C^3, such that no geodesic on M has on it two mutually conjugate points. Any unending geodesic on M is either a simple closed*

curve or the topological image of a line. The unending geodesics on M can be grouped into $\aleph$ families such that the members of any one family cover M, no two members of the same family intersect and any two members of different families do intersect.

The statements of this theorem follow readily from Theorems 5.4 and 5.7.

7. **Non-conjugacy and closed non-orientable surfaces of genus two.** Let M be a two-dimensional Riemannian manifold of the topological type of the Klein bottle (cf. Seifert-Threlfall [1, chap. 6]), or, as we shall term it, a closed non-orientable manifold of genus 2. (A Klein bottle is a closed non-orientable two-dimensional manifold with Euler characteristic 0.) The universal covering manifold of M can be mapped conformally onto Θ (cf. Koebe [1]) and we obtain a quadratic form (2.1) with $F(x, y)$ of class C^3 and invariant under a group G which, in this case, contains inversely conformal transformations of Θ into itself, as well as translations. The manifold M is obtained by identification of points which are congruent under G. The subset of transformations of G which are translations form a subgroup G^* which is generated by two linearly independent translations. The manifold M can be considered as obtained by identifying points congruent under G^*, thus obtaining a closed orientable manifold M^* of genus one, and then identifying the points of M^* in pairs under an inversely conformal transformation. The statements of the following theorem are then obtained with the aid of Theorems 5.4 and 6.1.

THEOREM 7.1. *Let M be a non-orientable two-dimensional Riemannian manifold of genus two, of class C^3, and such that no geodesic on M has on it two mutually conjugate points. The unending geodesics on M can be grouped into $\aleph$ families, of which $\aleph_0$ are made up of periodic geodesics, such that at each point of M there are in general two directions (in special cases one direction) which determine members of the same family.*

8. **The non-focality hypothesis.** Let g be an unending geodesic on a doubly-periodic manifold $M(F, \text{I})$. Corresponding to an arbitrary geodesic g_1 orthogonal to g at a point of g the first focal point Q of g on g_1, taking g_1 in either sense, is well defined, or fails to exist (cf. Bolza [1, §39]). Each such point Q will be termed a *focal point* of g.

(III) THE NON-FOCALITY HYPOTHESIS. *The Riemannian manifold $M(F, \text{I}, \text{II})$ will be said to satisfy the non-focality hypothesis if there exists a periodic L-line with the following property. Corresponding to each unending geodesic g of the type of L there are no focal points of g on at least one side of g.*

We denote a manifold $M(F, \text{I}, \text{II})$ which satisfies the non-focality hypothesis by $M(F, \text{I}, \text{II}, \text{III})$.

LEMMA 8.1. *Let $M(F, \text{I}, \text{II}, \text{III})$ satisfy the non-focality hypothesis with respect to the periodic E-line L. Then if g is any unending geodesic on $M(F, \text{I}, \text{II}, \text{III})$*

of the type of L and P is any point of Θ, there exists a unique geodesic passing through P and orthogonal to g.

If P is on g the statement of the lemma is obvious.

If P is not on g, the distance $D(P, Q)$, Q any point on g, has a minimum and assumes this minimum for some point T of g. The geodesic g^* passing through P and T is orthogonal to g. Thus the existence of at least one geodesic g^* satisfying the desired condition is proved.

Since the non-conjugacy hypothesis is fulfilled, g is of class A, g is simple, and thus g divides the points of Θ not on g into two open sets Θ_1 and Θ_2 which are separated by g. The non-focality hypothesis implies that in at least one of the sets Θ_1, Θ_2, and we can assume it to be Θ_1, there is no point P with two geodesics passing through P and orthogonal to g.

Suppose g_1 and g_2 are both orthogonal to g at P_1 and P_2, respectively, and intersect in a point Q which is necessarily in Θ_2. Since L is periodic and g is of the type of L, it follows from Theorem 5.3 that g is periodic. Accordingly there exists a geodesic g_1^* which is congruent to g_1 and intersects g orthogonally at a point P_1^* such that P_2 lies on the open segment $P_1P_1^*$ of g. The subset of g_1 (g_1^*) in Θ_1 is a ray which we denote by r_1 (r_1^*). Let the subset of Θ_1 bounded by $P_1P_1^*$, r_1 and r_1^* be denoted by α. The geodesic g_2 enters α at P_2, g_2 can have no other point in common with g other than P_2, and g_2 cannot intersect r_1 or r_1^*, for this would imply the existence of different geodesics orthogonal to g and meeting in Θ_1. But then one of the rays into which P_2 divides g_2 is of the type of r_1 (or r_1^*). It follows that g_2 is of the type of g_1 and these geodesics cannot intersect in Q as was assumed.

The proof of the lemma is complete.

THEOREM 8.1. *The Gaussian curvature of an $M(F$, I, II, III$)$ vanishes identically.*

Let L be the periodic E-line with respect to which the non-focality hypothesis is fulfilled. All the geodesics on an $M(F$, I, II, III$)$ which are of the type of L are periodic (Theorem 5.3) and no two intersect (Theorem 5.5). If we denote this family by $\mathcal{J}$, we have seen (cf. Theorem 5.7 and proof) that the geodesics of $\mathcal{J}$ form a field in Θ.

Let g be any geodesic in the family $\mathcal{J}$. According to Lemma 8.1, the geodesics orthogonal to g also form a field $\mathcal{J}^*(g)$ in Θ. Since two class A geodesics which are of different types intersect, it follows that the geodesics in $\mathcal{J}^*(g)$ are of the same type and include all the geodesics of this type.

Let g_1 and g_2 be two geodesics in the family $\mathcal{J}$. If P_1 is any point of g_1 and P_2 is any point of g_2, $D(P_1, P_2)$ is defined and varies continuously with P_1 and P_2. Since g_1 and g_2 are periodic with the same period, $D(P_1, P_2)$ assumes all its values, if P_1 is restricted to a properly chosen finite segment ω of g_1. It follows that $D(P_1, P_2)$ assumes its minimum for some pair Q_1, Q_2, Q_1 in ω, Q_2 on g_2, and the geodesic segment ρ joining Q_1 and Q_2 is orthogonal to both

g_1 and g_2. The unending geodesic g^* of which ρ is a segment belongs to both the families $\mathcal{J}^*(g_1)$ and $\mathcal{J}^*(g_2)$ and hence these families are identical. Thus the family $\mathcal{J}^*(g)$ is uniquely determined by L, and we denote this family by $\mathcal{J}^*$.

Each of the families $\mathcal{J}$ and $\mathcal{J}^*$ forms a field in Θ and each member of $\mathcal{J}$ intersects each member of $\mathcal{J}^*$ orthogonally in a single point. It is well known that this implies that the Gaussian curvature $K(x, y)$ vanishes identically. If this were not the case, there would exist a point $(\bar{x}, \bar{y})$ such that $K(\bar{x}, \bar{y}) \neq 0$. The function $K(x, y)$ is continuous and consequently there would exist a circle ν with center $(\bar{x}, \bar{y})$ such that either $K(x, y) > 0$ or $K(x, y) < 0$ in ν. But by proper choice of two members of the field $\mathcal{J}$ and two members of the field $\mathcal{J}^*$ we obtain a geodesic quadrilateral q with angles all right angles and lying in ν. The Gauss-Bonnet formula implies

$$\iint_q KF^2 dx dy = 0,$$

and we have a contradiction. Thus $K(x, y) \equiv 0$ and the proof of Theorem 6.1 is complete.

If M is a closed two-dimensional Riemannian manifold of class C^3, corresponding to any point P on any geodesic g on M there exists a set of focal points determined by P and g. We term the totality of focal points corresponding to all the points of g the *focal points corresponding to* g. If there are no focal points corresponding to any geodesic on M, there are no two mutually conjugate points on any geodesic on M. If, in addition, M is of the topological type of a torus or a non-orientable surface of genus two, the universal covering surface of M can be mapped onto the plane Θ and is an $M(F, \mathrm{I}, \mathrm{II}, \mathrm{III})$. If we term a Riemannian manifold for which the Gaussian curvature vanishes identically *flat*, Theorem 8.1 implies the sufficiency in the following theorem. The necessity is obvious.

THEOREM 8.2. *A two-dimensional Riemannian manifold M of class C^3 which is either orientable of genus one or non-orientable of genus two is flat, if and only if there are no focal points corresponding to any geodesic on M.*

PART II. MANIFOLDS OF HYPERBOLIC TYPE

9. A class of simply-connected Riemannian manifolds of hyperbolic type. Let U be the unit circle $u^2+v^2=1$ and let Ψ be its interior. To Ψ we assign the metric

$$ds^2 = \frac{4f^2(u, v)(du^2 + dv^2)}{(1 - u^2 - v^2)^2} \tag{9.1}$$

where $f(u, v)$ is of class C^3 in Ψ and there exist positive constants a and b such that

$$a \leqq f(u, v) \leqq b, \qquad u^2 + v^2 < 1. \tag{9.2}$$

We denote this two-dimensional Riemannian manifold by $M(f)$ and term it a *Riemannian manifold of hyperbolic type.*

The metric (9.1) assigns a length to any rectifiable curve γ on $M(f)$ and we denote this length by $L(\gamma)$. The geodesics corresponding to (9.1) are of class C^2 in terms of arc length as parameter and in terms of initial conditions. There is a unique geodesic passing through a point of Ψ in a given direction.

In the special case when $f(u, v) \equiv 1$, (9.1) reduces to the well known Poincaré metric which defines a hyperbolic geometry in Ψ. In this case the geodesics are arcs of circles orthogonal to U and will be termed *hyperbolic lines* or *H-lines*. The *hyperbolic length* or *H-length* of a curve in Ψ is the length of the curve under the assumption $f \equiv 1$. Given two points P and Q of Ψ, there is a unique H-line segment joining these points and the H-length of this segment is the *H-distance* $H(P, Q)$ between P and Q.

Again we term a geodesic segment g joining P and Q of *class A* if g affords an absolute minimum of length relative to all rectifiable curves on $M(f)$ joining P and Q. The manifold $M(f)$ is complete in the sense of Hopf and Rinow, and corresponding to a given pair of points P and Q of Ψ there exists a class A geodesic segment joining P and Q. We term the length of a class A geodesic segment joining P and Q the *distance* between P and Q and denote it by $D(P, Q)$. This metric has the usual properties.

A geodesic ray or an unending geodesic is of *class A*, if every finite segment of the ray or unending geodesic, respectively, is of class A. Geodesic segments geodesic rays, or unending geodesics of class A are evidently simple curves.

LEMMA 9.1. *If P and Q are arbitrary points of $M(f)$,*

$$aH(P, Q) \leqq D(P, Q) \leqq bH(P, Q).$$

For let γ be a class A geodesic segment joining P and Q. We then have

$$\begin{aligned} D(P, Q) = L(\gamma) &= \int_\gamma \frac{2f(\dot{u}^2 + \dot{v}^2)^{1/2}}{1 - u^2 - v^2}\, dt \geqq a \int_\gamma \frac{2(\dot{u}^2 + \dot{v}^2)^{1/2}}{1 - u^2 - v^2}\, dt \\ &= aH(\gamma) \geqq aH(P, Q). \end{aligned}$$

If we let h denote the H-line segment joining P and Q, we have

$$\begin{aligned} D(P, Q) = L(\gamma) \leqq L(h) &= \int_h \frac{2f(\dot{u}^2 + \dot{v}^2)^{1/2}}{1 - u^2 - \dot{v}^2}\, dt \\ &\leqq b \int_h \frac{2(\dot{u}^2 + \dot{v}^2)^{1/2}}{1 - u^2 - v^2}\, dt = bH(P, Q). \end{aligned}$$

The proof of the lemma is complete.

Let X and Y be point sets of Ψ. If P is any point of Ψ we define the *distance from the point P to the set X*, or $D(P, X)$, by

$$D(P, X) = \underset{x \in X}{\text{g.l.b.}}\ D(P, x).$$

We define the *type-distance* between the sets X and Y, or $\mathcal{D}(X, Y)$, by

$$\mathcal{D}(X, Y) = \max\left(\underset{x \in X}{\text{l.u.b.}}\ D(x, Y);\ \underset{y \in Y}{\text{l.u.b.}}\ D(y, X)\right).$$

Two sets will be said to be *of the same type* if their type-distance is finite.

THEOREM 9.1. *There exists a finite constant R, determined by $f(u, v)$ such that the type-distance between any class A geodesic segment of $M(f)$ and the H-line segment with the same end points never exceeds R.*

This theorem is essentially proved in Morse [1]. The proof follows from Lemma 9.1 of the present paper and Lemmas 4, 5, 6, 7, 8 of Morse [1].

We can state the following theorems (cf. Morse [2, p. 54]).

THEOREM 9.2. *Every class A unending geodesic on an $M(f)$ is of the type of some H-line. Conversely, corresponding to an arbitrary H-line h, there exists at least one unending class A geodesic on $M(f)$ of the type of h. The type-distance between an unending class A geodesic and an H-line of the same type cannot exceed the constant R of Theorem* 9.1.

THEOREM 9.3. *On an $M(f)$ a class A geodesic ray having P as initial point is of the type of a unique H-ray having P as initial point. Conversely, corresponding to an H-ray with initial point P, there exists a class A geodesic ray on $M(f)$ with initial point P and of the type of the H-ray. The type-distance between two geodesic rays of class A of the same type and with the same initial point, or between a geodesic ray of class A and an H-ray of the same type and with the same initial point cannot exceed R.*

An H-line determines and is determined by its two points at infinity on U. It follows from Theorem 9.2 that an unending geodesic g of class A on an $M(f)$ has among its limit points (in the euclidean sense) two and only two points of U. We term these the *points at infinity* of g.

10. **A class of two-dimensional Riemannian manifolds of hyperbolic type.** A one-to-one analytic transformation of Ψ into itself which leaves the metric

$$ds^2 = \frac{4(du^2 + dv^2)}{(1 - u^2 - v^2)^2} \tag{10.1}$$

invariant is conformal and hence is either a linear fractional transformation of the form (cf. Ford [1, chap. 1])

$$w = \frac{az + \bar{c}}{cz + \bar{a}}, \qquad a\bar{a} - c\bar{c} = 1,$$

or else is the product of such a transformation and an inversion in a circle

orthogonal to U. Let G be a properly discontinuous group of these transformations. That is, G fulfills the condition that the set of points congruent under G to any point P of Ψ does not have P as limit point. Groups such as G are said to be of the *first* or *second kind* according as all or not all points of U are limit points of fixed points of transformations of the group.

If $f(u, v)$ of (9.1) is invariant under the transformations of G, the metric (9.1) is also invariant under these same transformations, and if congruent points are identified there is defined a manifold $M(f, G)$ with properties depending on the properties of f and G. A geodesic on $M(f, G)$ is *represented* in Ψ by a finite or denumerable set of congruent geodesics of $M(f)$.

Let E denote the set of elements (u, v, ϕ), where $u^2+v^2<1$ and $0\leq\phi<2\pi$. The point (u, v, ϕ) of E determines a point $P(u, v)$ of Ψ and a direction ϕ at P, where we assume that ϕ is measured from a direction parallel to the positive u-axis. The point $P(u, v)$ is the point *bearing* e. To topologize the set E we consider E as the topological product of Ψ and a circle with central angle ϕ.

A transformation of G transforms a point of E into a congruent point or element and we denote by Ω the space obtained by identifying congruent points of E. The space Ω is the space of elements on $M(f, G)$. A point of Ω is represented in Ψ by a finite or infinite set of congruent elements.

We define measure in Ω by means of the integral

$$\iiint \frac{4f^2(u, v)}{(1 - u^2 - v^2)^2}\, du\,dv\,d\phi.$$

It is well known that this measure is invariant under the flow T_s defined in Ω by the directed geodesics on $M(f, G)$ with s the arc length on the geodesics.

Let p be a point of Ω. The set T_sp, $-\infty<s<+\infty$, will be termed the *motion* determined by p. A motion is the set of elements of Ω on a directed geodesic on $M(f, G)$. The motion determined by p is *stable in the sense of Poisson* if there exists an infinite sequence $\cdots <s_{-1}<s_0<s_1<s_2< \cdots$ with $\lim_{|n|\to+\infty} |s_n| = +\infty$, such that $\lim_{|n|\to+\infty} T_{s_n}p=p$.

DEFINITION. *The manifold $M(f, G)$ will be said to be Poisson stable provided almost all points of Ω are on motions which are stable in the sense of Poisson.* (Compare E. Hopf [2, p. 271].)

The class of manifolds $M(f, G)$ which are Poisson stable includes all those for which Ω is of finite measure, and hence those manifolds $M(f, G)$ which are of finite area. In particular, if G is a Fuchsian group of the first kind with a finite set of generators, $M(f, G)$ is Poisson stable.

If f is invariant under the group G and $M(f, G)$ is Poisson stable, we say that $M(f)$ is *Poisson stable* and denote it by $M(f, \mathrm{I})$.

11. The non-conjugacy hypothesis. We make the following definition.

(II) THE NON-CONJUGACY HYPOTHESIS. *There is no pair of mutually conjugate points on any geodesic on $M(f)$.*

The proofs of the following two theorems are similar to those of Theorems 5.1 and 5.2.

THEOREM 11.1. *A necessary and sufficient condition that $M(f)$ fulfill the non-conjugacy hypothesis is that there be only one geodesic segment joining two given points of $M(f)$.*

THEOREM 11.2. *A necessary and sufficient condition that $M(f)$ fulfill the non-conjugacy hypothesis is that all geodesics on $M(f)$ be of class A.*

If all the geodesics on $M(f)$ are of class A, it follows from Theorem 9.2 that each unending geodesic is of the type of an H-line, but we cannot conclude that there is only one geodesic of the type of a given H-line. Under the same condition, each geodesic ray on $M(f)$ is of class A and it follows from Theorem 9.3 that a geodesic ray issuing from a point P of Ψ is of the type of an H-ray issuing from P. It is conceivable however that there might be an infinite number of geodesic rays issuing from P and of the type of the same H-ray.

We denote a manifold $M(f, \mathrm{I})$ which satisfies the non-conjugacy hypothesis by $M(f, \mathrm{I}, \mathrm{II})$.

12. **Properties of the geodesics on an** $M(f, \mathrm{I}, \mathrm{II})$. We first prove the following lemma.

LEMMA 12.1. *Let P be a point of Ψ and let A, on U, be a fixed point of a hyperbolic transformation of G. If all the geodesics on $M(f)$ are of class A, there is but one geodesic ray from P to A.*

For suppose that there are two geodesic rays r_1 and r_2 from P to A. It follows from a theorem of Morse (cf. Morse [1, Theorem 10]) that each of the geodesic rays issuing from P and with A as point at infinity is either periodic or asymptotic to a periodic geodesic and all such periodic geodesics are of the same type. No two of the geodesic rays from P to A can be asymptotic to the same periodic geodesic (cf. Morse [1, Theorem 6]). Let r_3 be a geodesic ray issuing from P into the open set Π of Ψ bounded by r_1 and r_2. Since all geodesics are of class A, r_3 cannot cross r_1 or r_2 in any other point than P and r_3 must have A as point at infinity. It follows that r_3 is either part of a periodic geodesic g or is asymptotic to a periodic geodesic g. Since r_3 cannot be asymptotic to either r_1 or r_2 (cf. Morse [1, Theorem 6]) g must have on it points in the domain Π. But g must emerge from Π and hence must cross either r_1 or r_2. Suppose that g and r_1 cross. If g_1 is the unending geodesic of which r_1 is part and if g_1 is periodic, then g and g_1 are class A periodic geodesics of the same type which cross and this is impossible (cf. Morse [1, Theorem 9]). If g_1 is not periodic it is asymptotic, when traced out in the sense approaching A, to a class A periodic geodesic b which is of the type of g. But now we have a class A geodesic g_1 which is asymptotic to a class A peri-

odic geodesic b and crosses a class A periodic geodesic g which is of the type of b. This is likewise impossible (cf. Morse [1, Lemma 11]).

Similar arguments hold if g crosses r_2. The proof of the lemma is complete.

THEOREM 12.1. *Two geodesic rays on an $M(f, \text{I}, \text{II})$ with the same initial point in Ψ cannot be of the same type.*

For suppose the theorem is not true. There would then be a point P of Ψ with two directed geodesic rays r_1 and r_2 with initial point P and of the same type. Since r_1 and r_2 are of the same type, they have the same point at infinity A. Let g_1 and g_2 be the directed geodesics in Ψ of which r_1 and r_2, respectively, are rays. Let B_1 and B_2 be the points at infinity, different from A, of the geodesics g_1 and g_2, respectively. We do not assume that B_1 and B_2 are necessarily distinct.

Since all geodesics are of class A, the rays r_1 and r_2 together with their common point at infinity A form a Jordan curve which bounds a region κ' of Ψ. Similarly, the rays PB_1 and PB_2 of g_1 and g_2, respectively, and the arc B_1B_2 of U which does not contain A bound a region κ of Ψ. Let Q be any point of κ and Q' any point of κ'. The unique geodesic segment QQ' must cross both g_1 and g_2. It follows that the directed unending geodesic $g(Q, Q')$ of which the directed geodesic segment QQ' is part must have A as one of its points at infinity and its other point at infinity is in the arc B_1B_2. Let J denote the totality of directed geodesics $g(Q, Q')$, Q in κ, Q' in κ'.

In the space Ω corresponding to the manifold obtained by identifying congruent points of $M(f, \text{I}, \text{II})$ under the group G, the set J determines an open set J^*. But an open set is a measurable set of positive measure and it follows from the hypothesis that $M(f, \text{I}, \text{II})$ is Poisson stable that the set J^* contains a motion m which is stable in the sense of Poisson. Let g be a directed geodesic in the set J such that g represents m. Then A is one of the points at infinity of g and the other is some point B of the arc B_1B_2. Let Q be a point of κ on g, let r denote the directed geodesic ray QA which is part of g and let e be the initial element of r. It follows from the Poisson stability of m that there is a sequence of elements $e_1, e_2, \cdots$ on r such that if N_i is the point bearing e_i, then $\lim_{i\to+\infty} N_i = A$ (in the euclidean sense) and there exist elements $e_1', e_2', \cdots$, respectively congruent to $e_1, e_2, \cdots$, such that $\lim_{n\to+\infty} e_n' = e$.

Let T_n be the transformation of G such that $T_n(e_n) = e_n'$. Since the set J^* is open, there exists an integer K such that e_n' determines a point in J^* if $n > K$. Since all the directed geodesic rays with initial element determining a point in the set J^* can be represented in Ψ by rays with A as point at infinity, we infer that $T_n(A) = A$, $n > K$. We evidently have $\lim_{n\to+\infty} T_n(Q) = B$ (in the euclidean sense).

Since $T_n(A) = A$, $n > K$, T_n for $n > K$ is either: (1) a hyperbolic transformation with A as one of its fixed points; (2) a parabolic transformation with A as fixed point; (3) an inversion in a circle orthogonal to U and inter-

secting U in A; or (4) an inverse conformal transformation with A as one of its fixed points, the square of which is a hyperbolic transformation. But for n large, $T_n(Q)$ is near B and it is evident that (2) and (3) are excluded for n sufficiently large. Thus there exists a hyperbolic transformation T of G with A as one of its fixed points. According to Lemma 12.1 this is impossible if the rays r_1 and r_2 exist.

The proof of Theorem 12.1 is complete.

13. **Existence of transitive geodesics.** A directed unending geodesic (directed geodesic ray) on $M(f, G)$ is termed transitive if the elements on the geodesic (geodesic ray) form a set which is everywhere dense in the space Ω of elements on $M(f, G)$.

With a manifold $M(f, \mathrm{I}, \mathrm{II})$ there is associated a group G and a manifold $M(f, G)$ obtained from $M(f, \mathrm{I}, \mathrm{II})$ by identification of congruent points. A directed unending geodesic (directed geodesic ray) on $M(f, \mathrm{I}, \mathrm{II})$ is termed transitive if the corresponding geodesic (geodesic ray) on $M(f, G)$ is transitive. A necessary and sufficient condition that the directed geodesic g (geodesic ray r) on $M(f, \mathrm{I}, \mathrm{II})$ be transitive is that the totality of elements on g (r) and all its congruent copies form a set which is everywhere dense in the set E.

LEMMA 13.1. *If there exists a point P of Ψ such that for I, an arbitrary interval of U, the totality of elements on all the geodesic rays of $M(f, \mathrm{I}, \mathrm{II})$ from P to points of I and on all copies of these rays forms a set which is dense in E, there exists a transitive geodesic ray from P to some point of I.*

The space E contains a fundamental sequence $V_1, V_2, \cdots$ of neighborhoods such that any neighborhood of any point of E contains a neighborhood of this sequence. It follows from the hypothesis of the lemma that there must exist a closed interval I_1 of I such that any geodesic ray r_1 from P to I_1 either has on it, or one of the copies of r_1 has on it, an element in V_1. There exists a closed subinterval I_2 of I_1 which has analogous properties with respect to V_2. Continuing thus, we obtain a sequence of closed intervals $I_1 \supset I_2 \supset I_3 \supset \cdots$ and these must have a point A in common. The geodesic ray from P to A is evidently transitive.

LEMMA 13.2. *The group G corresponding to an $M(f, \mathrm{I}, \mathrm{II})$ is of the first kind.*

The group G is either of the first or second kind. If G is of the second kind it is known (cf. Ford [1, p. 73]) that the fundamental region of G abuts on the unit circle U in at least one interval I. But any geodesic on $M(f, G)$ which is represented in Ψ by a geodesic with a point at infinity in the interior of I is not stable in the sense of Poisson. Since the set of elements on all such directed geodesics forms an open set in Ω we conclude that $M(f, \mathrm{I}, \mathrm{II})$ is not Poisson stable, contrary to hypothesis.

THEOREM 13.1. *There exist transitive directed geodesic rays on any $M(f, \mathrm{I}, \mathrm{II})$.*

The proof of this theorem is similar to that of Theorem 3.1 of Hedlund [2]. We include a proof here for completeness.

It follows from Lemma 13.1 that it is sufficient to show that, if P is a point of Ψ and I is an arbitrary interval of U, then the set $e(P, I)$ consisting of the elements on all the geodesic rays from P to I and on all their copies is everywhere dense in E.

Let $e(x, y, \phi)$ be an arbitrary point of E. The element e determines a directed geodesic g, namely, that directed geodesic which passes through $S(x, y)$ the point bearing e and has the direction ϕ at S. Let A be the initial point at infinity of g, B the terminal point at infinity of g, and h the directed H-line AB. It follows from Theorem 9.2 that there is a point S' on h such that $D(S, S')$ is not greater than the constant R.

According to known results (cf. Koebe [1, (II), p. 349]) concerning the existence of transitive hyperbolic rays, it follows from the fact that G is of the first kind (Lemma 13.2) that there exists a directed hyperbolic ray PC, C an interior point of I, such that the directed hyperbolic ray PC is transitive. That is, the elements on PC together with the elements on all copies of PC form a set which is everywhere dense in E. It follows that there exists a sequence T_i, $i=1, 2, \cdots$, of transformations of G such that (in the euclidean sense) $\lim_{n\to+\infty} T_n(P)=A$ and $\lim_{n\to+\infty} T_n(C)=B$. Let $P_n=T_n(P)$ and consider the sequence of directed geodesic segments P_nS. With increasing n the direction of P_nS at S must approach that of g at S. For otherwise a subsequence of the set of geodesic segments P_nS, $n=1, 2, \cdots$, could be chosen such that this subsequence has as limiting geodesic ray a geodesic ray r with finite end point S, with point at infinity A, and such that r is not identical with the part AS of g. According to Theorem 12.1 this cannot be the case. Thus, if e_n is the element of P_nS at S, we infer that $\lim_{n\to+\infty} e_n=e$.

To complete the proof of the theorem it remains to show that for n sufficiently large, there is an element e_n', congruent to e_n, in the set $e(P, I)$. The points $T_n^{-1}(S')$ have the property that (in the euclidean sense) $\lim_{n\to+\infty} T_n^{-1}(S')=C$. The point $T_n^{-1}(S)$ is at a distance not exceeding R from $T_n^{-1}(S')$ and hence (in the euclidean sense) $\lim_{n\to+\infty} T_n^{-1}(S)=C$. There exists an N such that for $n>N$, all the points $T_n^{-1}(S)$ lie in the domain bounded by I and the geodesic rays from P to the end points of I. But then $T_n^{-1}(P_nS)$, $n>N$, is a segment of one of the geodesic rays from P to I, and if we let $e_n'=T_n^{-1}(e_n)$, the element e_n' is congruent to e_n and in the set $e(P, I)$.

The proof of the theorem is complete.

A directed geodesic g on a two-dimensional Riemannian manifold M is transitive, if the elements on g form a set which is everywhere dense among the totality of elements on M.

COROLLARY 13.1. *There exist transitive geodesics on any closed orientable two-dimensional Riemannian manifold M of class C^3 and of genus greater than*

one provided no geodesic on M has on it two mutually conjugate points. There exist transitive geodesics on any closed non-orientable two-dimensional Riemannian manifold M of class C^3 and of genus greater than two provided no geodesic on M has on it two mutually conjugate points.

In either case, the universal covering surface of M can be mapped conformally onto the interior Ψ of the unit circle, thus determining a manifold $M(f, \mathrm{I}, \mathrm{II})$ and a group G. The manifold M is obtained by identification of points which are congruent under G and the stated corollary follows from Theorem 13.1.

Bibliography

O. Bolza

1. *Vorlesungen über Variationsrechnung*, Berlin, 1909.

C. Carathéodory

1. *Über den Wiederkehrsatz von Poincaré*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, 1919, pp. 580–584.

L. R. Ford

1. *Automorphic Functions*, New York, 1929.

F. Hausdorff

1. *Mengenlehre*, Leipzig, 1927.

G. A. Hedlund

1. *Geodesics on a two-dimensional Riemannian manifold with periodic coefficients*, Annals of Mathematics, (2), vol. 33 (1932), pp. 719–739.
2. *Two-dimensional manifolds and transitivity*, Annals of Mathematics, (2), vol. 37 (1936), pp. 534–542.
3. *The dynamics of geodesic flows*, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 241–260.

E. Hopf

1. *Zwei Sätze über den wahrscheinlichen Verlauf der Bewegungen dynamischer Systeme.* Mathematische Annalen, vol. 103 (1930), pp. 710–719.
2. *Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung*, Berichten der Sächsischen Akademie der Wissenschaften zu Leipzig, 1939, pp. 261–304.

H. Hopf and W. Rinow

1. *Ueber den Begriff der vollständigen differentialgeometrischen Fläche*, Commentarii Mathematici Helvetici, vol. 3 (1931), pp. 209–225.

P. Koebe

1. *Riemannsche Mannigfaltigkeiten und nicht euklidische Raumformen*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, (I), 1927, pp. 164–196; (II), 1928, pp. 345–442.

M. Morse

1. *A fundamental class of geodesics on any closed surface of genus greater than one*, these Transactions, vol. 26 (1924), pp. 25–60.
2. *Instability and transitivity*, Journal de Mathématiques Pures et Appliquées, (9), vol. 14 (1935), pp. 49–71.

S. B. Myers

1. *Riemannian manifolds in the large*, Duke Mathematical Journal, vol. 1 (1935), pp. 39–49.

H. Poincaré

1. *Les Méthodes Nouvelles de la Mécanique Céleste*, vol. 3, Paris, 1899.

W. Rinow

1. See H. Hopf.
2. *Ueber Zusammenhänge zwischen der Differentialgeometrie im Grossen und im Kleinen*, Mathematische Zeitschrift, vol. 35 (1932), p. 522.

H. Seifert and W. Threlfall

1. *Lehrbuch der Topologie*, Leipzig, 1934.

Institute for Advanced Study,
Princeton, N. J.
University of Virginia,
Charlottesville, Va.

COMPLETELY CONVEX FUNCTIONS AND LIDSTONE SERIES

BY

D. V. WIDDER

Introduction. R. P. Boas suggested to the author the possibility of using Lidstone series to discuss the analytic character of functions whose even derivatives are positive on an interval. The author has described in an earlier note(1) how this suggestion led him to prove that a function for which

$$(1) \qquad (-1)^n f^{(2n)}(x) \geqq 0 \qquad (n = 0, 1, 2, \cdots)$$

on an interval is necessarily entire. In that note it was stated that we would show later the relation of this result to the problem of the representation of functions by Lidstone series. It is the aim of the present paper to study this relation.

A Lidstone series is a generalization of Taylor's series. It approximates to a given function in the neighborhood of two points instead of one. Such series have been studied by G. J. Lidstone(2), H. Poritsky(3), J. M. Whittaker(4), I. J. Schoenberg(5) and others. More precisely the series has the form

$$(2) \quad f(x) = f(1)\Lambda_0(x) + f(0)\Lambda_0(1-x) + f''(1)\Lambda_1(x) + f''(0)\Lambda_1(1-x) + \cdots,$$

where $\Lambda_n(x)$ is a polynomial of degree $2n+1$ defined by the relations

$$\Lambda_0(x) = x,$$
$$\Lambda_n''(x) = \Lambda_{n-1}(x),$$
$$\Lambda_n(0) = \Lambda_n(1) = 0 \qquad (n = 1, 2, \cdots).$$

Thus it is clear that the sum of an even number of terms of the series (2) is a polynomial which coincides with $f(x)$ at $x=0$ and at $x=1$. Moreover, each even derivative of the polynomial which does not vanish identically coincides with the corresponding derivative of $f(x)$ at those points.

Presented to the Society, January 1, 1941, under the title *Necessary and sufficient conditions for the representation of Lidstone series*; received by the editors February 21, 1941.

(1) D. V. Widder, *Functions whose even derivatives have a prescribed sign*, Proceedings of the National Academy of Sciences, vol. 26 (1940), pp. 657–659.

(2) G. J. Lidstone, *Notes on the extension of Aitken's theorem (for polynomial interpolation) to the Everett types*, Proceedings of the Edinburgh Mathematical Society, (2), vol. 2 (1929), pp. 16–19.

(3) H. Poritsky, *On certain polynomial and other approximations to analytic functions*, these Transactions, vol. 34 (1932), pp. 274–331.

(4) J. M. Whittaker, *On Lidstone's series and two-point expansions of analytic functions*. Proceedings of the London Mathematical Society, vol. 36 (1934), pp. 451–469.

(5) I. J. Schoenberg, *On certain two-point expansions of integral functions of exponential type*, Bulletin of the American Mathematical Society, vol. 42 (1936), pp. 284–288.

Boas has suggested, following the analogy with the completely monotonic functions of S. Bernstein, that we call functions satisfying (1) *completely convex*. We adopt this terminology and show that any function which is completely convex in an interval of length greater than π has an expansion (2) with every term positive. Since the sum of a Lidstone series must be entire we have here a new proof of the result mentioned above.

This sufficient condition for expansion in Lidstone series is not necessary, as simple examples show. In earlier papers, necessary conditions and sufficient conditions have been obtained, but none that are both necessary and sufficient. In the present paper such conditions are obtained by introducing the class of *minimal* completely convex functions, a slight variation of the class described above. It is proved, in fact, that a real function can be expanded in an absolutely convergent Lidstone series if and only if it is the difference of two minimal completely convex functions.

1. **The Green's function of a certain differential system.** Define a function $G(x, t)$ as follows

$$\begin{aligned} G(x, t) &= (x-1)t && (0 \leqq t < x \leqq 1), \\ &= (t-1)x && (0 \leqq x \leqq t \leqq 1). \end{aligned}$$

If $\phi(x)$ is any function continuous in the interval $0 \leqq x \leqq 1$, then it is easily verified that the unique solution of the differential system,

$$\begin{aligned} f''(x) &= \phi(x), \\ f(0) &= f(1) = 0, \end{aligned} \tag{1.1}$$

is

$$f(x) = \int_0^1 G(x, t)\phi(t)dt. \tag{1.2}$$

Consider next the successive iterates of $G(x, t)$, defined by the equations

$$\begin{aligned} G_1(x, t) &= G(x, t), \\ G_n(x, t) &= \int_0^1 G(x, y)G_{n-1}(y, t)dy \qquad (n = 2, 3, \cdots). \end{aligned}$$

From the property of $G(x, t)$ described above it is now clear that the function,

$$f(x) = \int_0^1 G_n(x, t)\phi(t)dt,$$

is the unique solution of the differential system

$$\begin{aligned} f^{(2n)}(x) &= \phi(x), \\ f^{(2k)}(0) &= f^{(2k)}(1) = 0 \qquad (k = 0, 1, \cdots, n-1). \end{aligned}$$

In fact $G_n(x, t)$ is the familiar Green's function of the system. Observe that $(-1)^n G_n(x, t)$ is non-negative in the unit square $0 \leqq x \leqq 1$, $0 \leqq t \leqq 1$.

Finally we see that the unique solution of the system

$$f^{(2n)}(x) = \phi(x), \quad f^{(2k)}(0) = a_{2k}, \quad f^{(2k)}(1) = b_{2k} \quad (k = 0, 1, \cdots, n-1), \tag{1.3}$$

is

$$f(x) = a_0(1 - x) + b_0 x + \sum_{k=1}^{n-1} a_{2k} \int_0^1 G_k(x, t)(1 - t)dt$$
$$+ b_{2k} \int_0^1 G_k(x, t) t dt + R_n(x),$$

where

$$R_n(x) = \int_0^1 G_n(x, t)\phi(t) dt.$$

If in this equation we replace a_{2k}, b_{2k} and $\phi(x)$ by their values in terms of $f(x)$ as given by the differential system (1.3) we obtain an identity which holds for all functions $f(x)$ possessing a sufficient number of derivatives. In fact this identity can be obtained without any appeal to the theory of differential equations by integrating the last integral by parts. We state the result as a theorem.

THEOREM 1.1. *If $f(x)$ is a function of class C^{2n} in the interval $0 \leqq x \leqq 1$, then*

$$\begin{aligned} f(x) &= f(0)(1 - x) + f(1)x + \sum_{k=1}^{n-1} f^{(2k)}(0) \int_0^1 G_k(x, t)(1 - t)dt \\ &\quad + f^{(2k)}(1) \int_0^1 G_k(x, t) t\, dt + R_n(x), \end{aligned} \tag{1.4}$$

where

$$R_n(x) = \int_0^1 G_n(x, t) f^{(2n)}(t) dt.$$

2. **Certain Fourier expansions.** Starting with the familiar Fourier series

$$1 = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{\sin (2k + 1)\pi t}{(2k + 1)} \qquad (0 < t < 1),$$

we multiply the series by $G_1(x, t)$ and integrate with respect to t from zero to unity. By (1.1) and (1.2) we have

$$\int_0^1 G_1(x, t) \sin (2k + 1)\pi t \, dt = -\frac{\sin (2k + 1)\pi x}{(2k + 1)^2 \pi^2},$$

so that we obtain at once the Fourier expansion of $\int_0^1 G_1(x, t)dt$. Repeating the process n times we obtain

THEOREM 2.1. *For $0 \leqq x \leqq 1$ and $n = 1, 2, \cdots$ the following expansion holds:*

$$\int_0^1 G_n(x, t)dt = (-1)^n \frac{4}{\pi^{2n+1}} \sum_{k=0}^{\infty} \frac{\sin (2k+1)\pi x}{(2k+1)^{2n+1}}. \tag{2.1}$$

A result of a similar nature is obtained from the familiar series

$$t = \frac{2}{\pi} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \sin k\pi t \qquad (0 < t < 1).$$

We record the result in

THEOREM 2.2. *For $0 \leqq x \leqq 1$ and $n = 1, 2, \cdots$ the following expansion holds:*

$$\int_0^1 G_n(x, t)t\, dt = (-1)^n \frac{2}{\pi^{2n+1}} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^{2n+1}} \sin k\pi x. \tag{2.2}$$

3. **Lidstone polynomials.** The function (2.2) is clearly a polynomial of degree $2n+1$ since its derivative of order $2n$ is the function x. We shall refer to it as the Lidstone polynomial of order n and, following the original notation of Lidstone, set

$$\Lambda_0(x) = x,$$
$$\Lambda_n(x) = \int_0^1 G_n(x, t)t\, dt \qquad (n = 1, 2, \cdots).$$

By subtracting equation (2.2) from equation (2.1) we have

$$\int_0^1 G_n(x, t)(1-t)dt = (-1)^n \frac{2}{\pi^{2n+1}} \sum_{k=1}^{\infty} \frac{\sin k\pi x}{k^{2n+1}}.$$

But if we replace x by $(1-x)$ in this series we obtain the series (2.2), whose sum is $\Lambda_n(x)$. That is, the equation

$$\int_0^1 G_n(x, t)(1-t)dt = \Lambda_n(1-x)$$

holds at least for $0 \leqq x \leqq 1$. But the result is true for all x since both sides of the equation are polynomials. Equations (1.4) and (2.2) now become

$$f(x) = \sum_{k=0}^{n-1} [f^{(2k)}(0)\Lambda_k(x) + f^{(2k)}(1)\Lambda_k(1-x)] + \int_0^1 G_n(x, t) f^{(2n)}(t)dt, \tag{3.1}$$

$$\Lambda_n(x) = (-1)^n \frac{2}{\pi^{2n+1}} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^{2n+1}} \sin k\pi x \qquad (0 \leqq x \leqq 1). \tag{3.2}$$

4. **Asymptotic behavior of $\Lambda_n(x)$ for large n.** The first term of the Fourier expansion (3.2) serves as a close approximation to $\Lambda_n(x)$ if n is large. More explicitly we have

THEOREM 4.1. *There exists a constant M such that*

$$\left|(-1)^n\Lambda_n(x) - \frac{2}{\pi^{2n+1}}\sin \pi x\right| < \frac{M}{(2\pi)^{2n+1}} \quad (0 \leqq x \leqq 1; n = 0, 1, \cdots). \tag{4.1}$$

For, it is clear from equation (3.2) that for $0 \leqq x \leqq 1$,

$$\left|(-1)^n\Lambda_n(x) - \frac{2}{\pi^{2n+1}}\sin \pi x\right| \leqq \frac{2}{(2\pi)^{2n+1}}\left[1 + \left(\frac{2}{3}\right)^{2n+1} + \left(\frac{2}{4}\right)^{2n+1} + \cdots\right].$$

The result is now obvious since the Dirichlet series in brackets tends to unity when n becomes infinite.

THEOREM 4.2. *There exists a constant M such that*

$$0 \leqq (-1)^n\Lambda_n(x) \leqq \frac{M}{\pi^{2n}} \quad (0 \leqq x \leqq 1; n = 1, 2, \cdots).$$

This follows at once from (4.1).

THEOREM 4.3. *For any fixed x_0 between zero and unity there is a constant M such that*

$$(-1)^n\Lambda_n(x_0) \geqq \frac{M}{\pi^{2n}} \quad (n = 1, 2, \cdots).$$

For by (4.1) we see that

$$\lim_{n\to\infty} (-1)^n \frac{\Lambda_n(x_0)\pi^{2n}}{\sin \pi x_0} = \frac{2}{\pi} \quad (0 < x_0 < 1).$$

THEOREM 4.4. *There exists a constant M such that*

$$0 \leqq (-1)^n \int_0^1 G_n(x, t)dt \leqq \frac{M}{\pi^{2n}} \quad (0 \leqq x \leqq 1; n = 1, 2, \cdots).$$

This is proved from equation (2.1) as Theorem 4.1 was proved from equation (3.2).

5. **Lidstone series.** We denote the class of real entire functions of exponential type less than π by P. In order to make our results independent of the theory of entire functions we may introduce the class by means of the following definition.

DEFINITION 5.1. *A real entire function $f(x)$ belongs to the class P if there exists a positive number $p < \pi$ such that*

$$f^{(n)}(0) = O(p^n) \qquad (n \to \infty). \tag{5.1}$$

Concerning this class of functions we prove a preliminary result.

LEMMA 5.1. *If $f(x)$ belongs to P then there exists a positive number $p < \pi$ such that*

$$f^{(n)}(x) = O(p^n)$$

uniformly in $0 \leqq x \leqq 1$.

For, by Taylor's expansion

$$f^{(n)}(x) = \sum_{k=0}^{\infty} f^{(n+k)}(0) \frac{x^k}{k!}.$$

By (5.1)

$$|f^{(n)}(x)| \leqq M \sum_{k=0}^{\infty} p^{n+k} \frac{x^k}{k!} \leqq M p^n e^p \qquad (0 \leqq x \leqq 1),$$

where M is some constant. This proves the lemma. Note that if $f(x)$ belongs to P then $f(x+a)$ does also for every constant a. The interval (0, 1) of the lemma could be replaced by any finite interval.

THEOREM 5.1. *If $f(x)$ belongs to P, then*

$$f(x) = f(1)\Lambda_0(x) + f(0)\Lambda_0(1-x) + f''(1)\Lambda_1(x) + f''(0)\Lambda_1(1-x) + \cdots, \tag{5.2}$$

the series converging uniformly in $0 \leqq x \leqq 1$.

Denote by $s_n(x)$ the sum of the first n terms of the series (5.2). Then by Theorem 1.1, Theorem 4.4 and Lemma 5.2 there is a constant M such that

$$|f(x) - s_{2n}(x)| \leqq \int_0^1 (-1)^n G_n(x, t) |f^{(2n)}(t)| \, dt \leqq M \left(\frac{p}{\pi}\right)^{2n}$$
$$(n = 0, 1, \cdots).$$

Also

$$s_{2n+1}(x) = s_{2n}(x) + f^{(2n)}(0)\Lambda_n(x).$$

By Theorem 4.2 and Definition 5.1 we see that

$$\lim_{n \to \infty} f^{(2n)}(0)\Lambda_n(x) = 0$$

uniformly in $0 \leqq x \leqq 1$, so that the theorem is established.

THEOREM 5.2. *If the series*

$$b_0\Lambda_0(x) + a_0\Lambda_0(1-x) + b_1\Lambda_1(x) + a_1\Lambda_1(1-x) + \cdots \tag{5.3}$$

converges for a single value of x in $0 < x < 1$, it converges uniformly throughout that interval to a function $f(x)$. Then the series

$$(5.4) \qquad b_0 + a_0 - \frac{b_1}{\pi^2} - \frac{a_1}{\pi^2} + \frac{b_2}{\pi^4} + \frac{a_2}{\pi^4} - \cdots$$

converges and

$$(5.5) \quad f^{(2k)}(x) = b_k\Lambda_0(x) + a_k\Lambda_0(1-x) + b_{k+1}\Lambda_1(x) + a_{k+1}(1-x) + \cdots$$

for $0 \leqq x \leqq 1$ *and* $k = 0, 1, 2, \cdots$.

For, if the series (5.3) converges for $x = x_0$ then

$$\lim_{n\to\infty} b_n\Lambda_n(x_0) = 0, \qquad \lim_{n\to\infty} a_n\Lambda_n(1-x_0) = 0.$$

Then by Theorem 4.3

$$b_n = O(\pi^{2n}), \qquad a_n = O(\pi^{2n}).$$

This shows by (4.1) that the series

$$(5.6) \qquad \sum_{n=0}^{\infty} b_n\left[\Lambda_n(x_0) - (-1)^n\frac{2\sin \pi x_0}{\pi^{2n+1}}\right] + a_n\left[\Lambda_n(1-x_0) - (-1)^n\frac{2\sin\pi x_0}{\pi^{2n+1}}\right]$$

converges absolutely. Subtracting from this the series (5.3) convergent for $x = x_0$ we see that the resulting series

$$(5.7) \qquad -\frac{2\sin\pi x_0}{\pi}\sum_{n=0}^{\infty}(-1)^n\frac{b_n}{\pi^{2n}} + (-1)^n\frac{a_n}{\pi^{2n}}$$

must converge. That is, (5.4) must converge. When x_0 is replaced by the variable x in (5.6) and (5.7) it is clear that both series converge uniformly in $0 \leqq x \leqq 1$. The same must be true of their difference, the series (5.3).

Finally, to prove (5.5) we must prove the series (5.5) uniformly convergent in $0 \leqq x \leqq 1$. To see this we note that the series

$$(5.8) \qquad \sum_{n=0}^{\infty} b_{n+k}\left[\Lambda_n(x) - (-1)^n\frac{2\sin\pi x}{\pi^{2n+1}}\right] + a_{n+k}\left[\Lambda_n(1-x) - (-1)^n\frac{2\sin\pi x}{\pi^{2n+1}}\right],$$

$$(5.9) \qquad 2\sin\pi x\sum_{n=0}^{\infty}(-1)^n\frac{b_{n+k}}{\pi^{2n+1}} + (-1)^n\frac{a_{n+k}}{\pi^{2n+1}}$$

both converge uniformly in $0 \leqq x \leqq 1$ by Theorem 4.1 and by the convergence of (5.7). Hence the sum of the series (5.8) and (5.9), series (5.5), must also converge uniformly in $0 \leqq x \leqq 1$.

6. **Completely convex functions.** Since the successive Lidstone polynomials alternate in sign on the interval $0\leqq x\leqq 1$, it is natural to consider for expansion in Lidstone series functions whose even derivatives alternate in sign on that interval. We call such a function completely convex there.

DEFINITION 6.1. *A real function $f(x)$ is completely convex on the interval $a\leqq x\leqq b$ if it has derivatives of all orders there and if*

$$(-1)^k f^{(2k)}(x) \geqq 0 \qquad (a \leqq x \leqq b;\ k = 0, 1, 2, \cdots).$$

For example, the functions $\sin x$ and $\cos x$ are completely convex on the intervals $(0, \pi)$ and $(-\pi/2, \pi/2)$, respectively.

We now prove certain properties of functions of this class.

THEOREM 6.1. *If $f(x)$ is completely convex in $0\leqq x\leqq 1$, then*

$$\begin{aligned} f^{(2k)}(0) &= O(\pi^{2k}), \\ f^{(2k)}(1) &= O(\pi^{2k}) \qquad (k \to \infty). \end{aligned} \tag{6.1}$$

Consider the identity (3.1) for the present function $f(x)$. Since every term of the series is non-negative we have

$$\begin{aligned} &0 \leqq f^{(2k)}(0)\Lambda_k(x) \leqq f(x), \\ &0 \leqq f^{(2k)}(1)\Lambda_k(1-x) \leqq f(x) \qquad (0 \leqq x \leqq 1;\ k = 0, 1, \cdots). \end{aligned} \tag{6.2}$$

In particular, choose $x=1/2$ and apply Theorem 4.3. This gives the relations (6.1) at once.

THEOREM 6.2. *If $f(x)$ is completely convex in $0\leqq x\leqq 1$, then there is a constant M such that*

$$\begin{aligned} &0 \leqq (-1)^k f^{(2k)}(x) \leqq M\left(\frac{\pi}{x}\right)^{2k}, \\ &0 \leqq (-1)^k f^{(2k)}(x) \leqq M\left(\frac{\pi}{1-x}\right)^{2k} \qquad (k \to \infty). \end{aligned} \tag{6.3}$$

For, if $f(x)$ is completely convex in $a\leqq x\leqq b$, then the same is true of

$$F(x) = f(a + bx - ax)$$

in $0\leqq x\leqq 1$. By Theorem 6.1 we have for $0\leqq a<b\leqq 1$

$$\begin{aligned} F^{(2k)}(0) &= f^{(2k)}(a)(b-a)^{2k} = O(\pi^{2k}), \\ F^{(2k)}(1) &= f^{(2k)}(b)(b-a)^{2k} = O(\pi^{2k}) \qquad (k \to \infty). \end{aligned} \tag{6.4}$$

Choosing first $a=0$, $b=x<1$ and then $a=x>0$, $b=1$ we obtain (6.3) from (6.4). One sees easily from (6.2) that M is independent of x in (6.3).

We introduce next a familiar result of J. Hadamard[6] as

[6] See, for example, T. Carleman, *Les Fonctions Quasi-Analytiques*, Paris, 1912, p. 12.

LEMMA 6.1. *If $f(x)$ is of class C^2 in $a \leqq x \leqq b$ and if*

$$\max_{a \leqq x \leqq b} |f(x)| = M_0,$$

$$\max_{a \leqq x \leqq b} |f''(x)| = M_2$$

then

$$|f'(x)| \leqq \frac{2M_0}{b-a} + \frac{M_2(b-a)}{2} \qquad (a \leqq x \leqq b).$$

By use of this result we can now prove

THEOREM 6.3. *If $f(x)$ is completely convex in $a \leqq x \leqq b$ with $b-a>1$, then $f(x)$ belongs to class P and equation (5.2) holds.*

From (6.4) we have for a suitable constant M

$$|f^{(2k)}(x)| \leqq M\left(\frac{\pi}{b-x}\right)^{2k} \qquad (a \leqq x \leqq b).$$

Choose a number c so near a that $b-c>1$. Then

$$|f^{(2k)}(x)| \leqq M\left(\frac{\pi}{b-c}\right)^{2k} \qquad (a \leqq x \leqq c).$$

Setting $\pi/(b-c)=p$ we have by Lemma 6.1

$$|f^{(2k+1)}(x)| \leqq \frac{2M}{c-a}p^{2k} + M\frac{(c-a)}{2}p^{2k+2}.$$

That is,

$$f^{(k)}(x) = O(p^k) \qquad (k \to \infty),$$

uniformly in $(a \leqq x \leqq c)$. This shows that $f(x)$ is entire and that $f(x+a)$ belongs to P. It follows that $f(x)$ belongs to P and the theorem is proved.

7. **Minimal convex functions.** The sufficient condition of Theorem 6.3 for the representation of a function in Lidstone series is not necessary. For example, the function $\sinh x$ is not completely convex in any interval; yet it has the Lidstone expansion

$$\sinh x = \sinh 1 \sum_{n=0}^{\infty} \Lambda_n(x).$$

We shall see that a convergent Lidstone series with every term non-negative in $0 \leqq x \leqq 1$ defines a function which is completely convex there. But observe that $\sin \pi x$ has the same property. Yet it has no Lidstone representation since every term of the Lidstone series for this function is zero. To obtain conditions that are both necessary and sufficient we introduce a further definition.

DEFINITION 7.1. *A function $f(x)$ is a minimal completely convex function in the interval $0 \leqq x \leqq 1$ if it is completely convex there and if $f(x) - \epsilon \sin \pi x$ is not completely convex there for any positive ϵ.*

For example, the functions $f(x) = 0$ and $f(x) = \sin x$ are minimal convex functions in $0 \leqq x \leqq 1$ but the function $f(x) = \sin \pi x$ is not.

THEOREM 7.1. *If the series*

$$\sum_{n=0}^{\infty} (-1)^n b_n \Lambda_n(x) + (-1)^n a_n \Lambda_n(1-x), \tag{7.1}$$

$$a_n \geqq 0, \qquad b_n \geqq 0 \qquad (n = 0, 1, 2, \cdots)$$

converges to $f(x)$, then $f(x)$ is a minimal completely convex function in the interval $0 \leqq x \leqq 1$.

Differentiating series (7.1) we obtain

$$(-1)^k f^{(2k)}(x) = \sum_{n=0}^{\infty} (-1)^n b_{n+k} \Lambda_n(x) + (-1)^n a_{n+k} \Lambda_n(1-x).$$

Since the right-hand side is non-negative $f(x)$ is completely convex in $0 \leqq x \leqq 1$. By Theorem 4.2

$$(-1)^k f^{(2k)}(x) \leqq M \sum_{n=0}^{\infty} (a_{n+k} + b_{n+k}) \pi^{-2n} = M \pi^{2k} R_k,$$

where

$$R_k = \sum_{n=k}^{\infty} \frac{a_n}{\pi^{2n}} + \frac{b_n}{\pi^{2n}} \cdot$$

By Theorem 5.2, R_k is defined and tends to zero with $1/k$. For a given positive number ϵ and a number x_0 between zero and one we can find an integer k so large that

$$MR_k - \epsilon \sin \pi x_0 < 0.$$

That is, the function

$$(-1)^k [f(x) - \epsilon \sin \pi x]^{(2k)}$$

is negative at x_0. Hence $f(x)$ is a minimal completely convex function in $0 \leqq x \leqq 1$.

For our next result we need to prove a lemma.

LEMMA 7.1. *If $f(x)$ and $-f''(x)$ are non-negative in $0 \leqq x \leqq 1$ and if $f(x_0) > \epsilon\pi$ for some number x_0 in that interval, then*

$$f(x) > \epsilon \sin \pi x \qquad (0 \leqq x \leqq 1). \tag{7.2}$$

The result is obvious geometrically, but we give an analytical proof. If $0<x_0\leqq 1$ we have by the convexity of $f(x)$

$$f(x) \geqq f(0) + [\epsilon\pi - f(0)]x_0^{-1}x \qquad (0 \leqq x \leqq x_0),$$
$$f(x) \geqq [f(0)(x_0 - x) + \epsilon\pi x]x_0^{-1} \qquad (0 \leqq x \leqq x_0),$$
$$f(x) > \epsilon\pi x \geqq \epsilon \sin \pi x \qquad (0 \leqq x \leqq x_0).$$

If $x_0=1$ the proof is complete. Otherwise by applying what has just been proved to the function $f(1-x)$ we see that (7.2) also holds in $x_0\leqq x\leqq 1$. Finally, if $x_0=0$ the above proof is applicable to the function $f(1-x)$.

THEOREM 7.2. *If $f(x)$ is a minimal completely convex function in $0\leqq x\leqq 1$, then it can be expanded in a convergent Lidstone series.*

Define $s_n(x)$ as in §5. Then under the present hypothesis on $f(x)$ it is clear from equation (3.1) that

$$s_n(x) \leqq f(x) \qquad (0 \leqq x \leqq 1;\, n = 0, 1, \cdots),$$

and that $s_n(x)$ is a non-decreasing function of n for each x. Hence $s_n(x)$ tends to some function as n becomes infinite. We wish to prove that its limit is $f(x)$. Suppose the contrary and assume that for some x_0 in $0\leqq x\leqq 1$

$$f(x_0) - \lim_{n\to\infty} s_n(x_0) = \Delta > 0.$$

Then

$$f(x_0) - s_{2n}(x_0) = \int_0^1 G_n(x_0, t)f^{(2n)}(t)dt \geqq \Delta \qquad (n = 1, 2, \cdots). \tag{7.3}$$

Since $f(x)$ is a minimal completely convex function, $f(x)-\epsilon \sin \pi x$ fails to be completely convex in $0\leqq x\leqq 1$ for every positive ϵ. Choose $\epsilon<\Delta/(\pi M)$, where M is the constant of Theorem 4.4. Then there exists an integer k and a number x_0 in $(0, 1)$ such that

$$(-1)^k f^{(2k)}(x_0) - \epsilon\pi^{2k} \sin \pi x_0 < 0.$$

By virtue of Lemma 7.1 this implies that

$$(-1)^k f^{(2k)}(x) \leqq \epsilon\pi^{2k+1} \qquad (0 \leqq x \leqq 1),$$

so that by Theorem 4.4

$$\int_0^1 G_k(x_0, t)f^{(2k)}(t)dt \leqq \epsilon\pi M < \Delta.$$

This contradicts inequality (7.3). The assumption that $s_n(x)$ does not approach $f(x)$ as n becomes infinite is untenable, so that our theorem is proved.

8. **Necessary and sufficient conditions for representation.** We conclude by

proving a necessary and sufficient condition that $f(x)$ can be represented by an absolutely convergent Lidstone series

THEOREM 8.1. *A necessary and sufficient condition that $f(x)$ can be represented by an absolutely convergent Lidstone series is that it should be the difference of two minimal completely convex functions on* $0 \leqq x \leqq 1$.

To prove the sufficiency of the condition let

$$f(x) = g(x) - h(x),$$

where $g(x)$ and $h(x)$ are both minimal completely convex functions on $0 \leqq x \leqq 1$. By Theorem 7.2

$$g(x) = \sum_{n=0}^{\infty} g^{(2n)}(1)\Lambda_n(x) + g^{(2n)}(0)\Lambda_n(1 - x),$$

$$h(x) = \sum_{n=0}^{\infty} h^{(2n)}(1)\Lambda_n(x) + h^{(2n)}(0)\Lambda_n(1 - x).$$

Each series has only positive terms so that when we subtract them the result is an absolutely convergent Lidstone series whose sum is $f(x)$.

Conversely, assume that

$$f(x) = \sum_{n=0}^{\infty} b_n\Lambda_n(x) + a_n\Lambda_n(1 - x), \tag{8.1}$$

the series converging absolutely. Set

$$g(x) = \sum_{n=0}^{\infty} (-1)^n | b_n | \Lambda_n(x) + (-1)^n | a_n | \Lambda_n(1 - x),$$

$$\begin{aligned} h(x) = \sum_{n=0}^{\infty} & (-1)^n\{ | b_n | - (-1)^n b_n \}\Lambda_n(x) \\ & + (-1)^n\{ | a_n | - (-1)^n a_n \}\Lambda_n(1 - x). \end{aligned}$$

These series both converge since (8.1) converges absolutely. Every term of these two series is positive. Hence by Theorem 7.1 $g(x)$ and $h(x)$ are minimal completely convex. Since $f(x) = g(x) - h(x)$, our result is completely proved.

HARVARD UNIVERSITY
CAMBRIDGE, MASS.

ERGODIC THEOREMS FOR ABELIAN SEMI-GROUPS[1]

BY

MAHLON M. DAY

This paper adapts the method of F. Riesz to the proof of certain general ergodic theorems for Abelian semi-groups of operators on a Banach space to itself. The main features of the method are that no measurability conditions are imposed on the semi-group under consideration and that consistent use of the second conjugate space and its compactness properties make it possible to replace the compactness conditions often imposed by a more natural restriction on the transforms of points. Theorem 1 and the various supplementary results include as special cases theorems of Lorch [10], Dunford [7], Yosida [15], F. Riesz [12], and Cohen [6]. It overlaps the work of Alaoglu and Birkhoff [3] at those points where they consider Abelian cases; for example, Corollary 8 is a great generalization of their Theorem 5.

Section 1 contains some introductory material on conjugate spaces and adjoint operations. Section 2 introduces bounded Abelian semi-groups of operators and near invariance of a system of set functions on such a semi-group; this section also contains the principal theorem (Theorem 1) of the paper. The form of this theorem raises three questions ((A) to (C) at the beginning of §3). The answer to (A) shows, among other things, that every Abelian semi-group has a property much like "ergodicity" in the sense of Alaoglu and Birkhoff; Theorem 3 is the main result here. The answer to (B) again indicates the importance of reflexivity in theorems of this type; Corollary 8 is one example. Two special cases of (C) give a generalization of Dunford's theorem (Theorem 5) and a theorem on bounded Abelian semi-groups of projections (Theorem 6) which has not, so far as I know, been considered before.

1. **Some properties of Banach spaces.** If B is a Banach space[2], let B^* be the set of all linear—that is, additive and continuous—real-valued functions on B. If, for β in B^*, $\|\beta\| = \sup_{\|b\|\leq 1} |\beta(b)|$, then B^* is also a Banach space. As is usual, the weak neighborhood topology in B is defined as follows: For each b_0 in B the weak neighborhoods of b_0 are the sets of the form[3] $\{b \mid |\beta_i(b) - \beta_i(b_0)| < \epsilon \text{ for } i = 1, \cdots, k\}$ for every choice of $\epsilon > 0$, k a positive integer, and $\beta_1, \cdots, \beta_k$ in B^*. With this topology B is a linear topological

Presented to the Society in two parts, the first under the title *A norm ergodic theorem* on April 27, 1940; the second under the present title on September 5, 1941; received by the editors April 4, 1940, and, in revised form, March 21, 1941.

(1) Part of the work on this paper was done at the Institute for Advanced Study while the author held the Corinna Borden Keen Research Fellowship of Brown University.

(2) See Banach [4]. A Banach space is a complete norm vector space.

(3) $\{b \mid \cdots\}$ is the set of all points b satisfying the condition after the vertical bar.

space (see Wehausen [12]) and hence is a regular Hausdorff space in which addition of elements and multiplication by real numbers are continuous operations.

Since B^* is a Banach space, it has a weak topology defined just as in B; however, there is another topology in a conjugate space which cannot be defined in every space. This is the weak* topology in which the neighborhoods of a point β_0 in B^* are the sets of the form $\{\beta \mid |\beta(b_i)-\beta_0(b_i)|<\epsilon$ for $i=1, \cdots, k\}$ for all choices of $\epsilon>0$, k a positive integer, and $b_1, \cdots, b_k$ in B. Each weak* neighborhood of β_0 is a weak neighborhood of β_0 but the converse is not true unless B is reflexive. The importance of the weak* topology in this paper arises from

LEMMA 1. *The unit sphere in B^* is always compact*(4) *in the weak* neighborhood topology.*

This has been proved by Alaoglu [2], Kakutani [9], and Šmulian [13].

If B is a Banach space, let B^{**} be the space $(B^*)^*$. Then there is a natural imbedding of B in B^{**} which associates to each b in B the point $\mathfrak{b}_b$ in B^{**} such that $\mathfrak{b}_b(\beta)=\beta(b)$ for every β in B^*. B is *reflexive* if B fills up B^{**} under this imbedding. For the rest of this paper B will be considered to be imbedded in this way in B^{**} whenever it seems convenient.

If T is a linear operator defined on B with values in B, let T^*, the adjoint of T, be the operator on B^* to B^* such that $\beta(Tb)=T^*\beta(b)$ for every β in B^* and b in B. Then:

(1) $\|T^*\|=\|T\|$.

(2) $(T_1T_2)^*=T_2^*T_1^*$ so T_1^* and T_2^* commute if T_1 and T_2 do.

(3) If $T^{**}=(T^*)^*$, then T^{**} agrees with T in B; that is, $T^{**}\mathfrak{b}_b=\mathfrak{b}_{Tb}$, since $T^{**}\mathfrak{b}_b(\beta)=\mathfrak{b}_b(T^*\beta)=T^*\beta(b)=\beta(Tb)=\mathfrak{b}_{Tb}(\beta)$ for every β in B^*.

For brevity T will sometimes be used for T^{**}.

If Y is any set of elements y, M_Y is the Banach space of all real-valued bounded functions ϕ on Y with(5) $\|\phi\|_{M_Y}=\sup_{y\in Y}|\phi(y)|$. If B is any Banach space, $M_Y(B)$ is the Banach space of all bounded functions f on Y with values in B where $\|f\|_{M_Y(B)}=\sup_{y\in Y}\|f(y)\|_B$. If Υ is any element of M_Y^*, it is possible to define U on $M_Y(B)$ to B^{**} by letting $U(f)$ be that point $\mathfrak{b}$ of B^{**} such that $\mathfrak{b}(\beta)=\Upsilon(\beta f)$ for every β in B^*, where βf is the element of M_Y defined by $\beta f(y)=\beta(f(y))$. For each Υ in M_Y^* there is defined a unique, bounded, additive(6) set function Ψ by the relation $\Psi(E)=\Upsilon(\phi_E)$, where ϕ_E is the charac-

(4) E is compact (bicompact in the sense of Alexandroff and Hopf, *Topologie*, I) if every covering of E by open sets contains a finite subcovering; that is, if $E\subset\sum_\alpha O_\alpha$, where the O_α are open, there exist $\alpha_1, \cdots, \alpha_k$ such that $E\subset\sum_{i\leq k}O_{\alpha_i}$; this is equivalent to the following condition on closed sets: If the closed sets $C_\alpha\subset E$ are such that every finite set of the C_α have a point in common, then $\prod_\alpha C_\alpha$ is not empty.

(5) The subscript on the norm symbol indicates the space in question; it will be omitted when there is no danger of confusion.

(6) Ψ is bounded if $\Psi(E)\leq K$ if $E\subset Y$, Ψ is additive if $\Psi(E_1+E_2)=\Psi(E_1)+\Psi(E_2)$ whenever

teristic function of E. Conversely, for each bounded additive Ψ, setting $\mathfrak{T}(\phi) = \int \phi d\Psi$ defines an $\mathfrak{T}$ in M_Y^*, where the integral is, say, that of Radon-Stieltjes(7); moreover $\|\mathfrak{T}\| = V\Psi(Y)$. Because of this relation between Ψ and $\mathfrak{T}$ it is possible to define $\int f d\Psi$ for f in $M_Y(B)$ to be the element $\mathfrak{b}$ of B^{**} for which $\mathfrak{b}(\beta) = \int \beta f d\Psi$ for every β in B^*. All integrals used hereafter will be of this nature(8).

A set X is *directed* if there is a relation $>$ ("follows") among some pairs of its points such that $x > x'$ and $x' > x''$ implies $x > x''$ and such that each pair, x' and x'', of points in X has a common successor, x in X; that is, $x > x'$ and $x > x''$. If for each x in X, s_x is a point of the topological space S, then $s = \lim_x s_x$ if and only if for each neighborhood N of s there is an x_N in X such that $s_x \in N$ if $x > x_N$.

LEMMA 2. *If X is a directed set, if for each x in X $\mathfrak{b}_x$ is a point in B^{**} (B any Banach space), and if $\|\mathfrak{b}_x\|$ is ultimately bounded—that is, if there exist $K > 0$ and $x_0 \in X$ such that $\|\mathfrak{b}_x\| \leq K$ if $x > x_0$—then there is a $\mathfrak{b}_0$ in B^{**} such that $\mathfrak{b}_0$ is in the weak* closure of $\{\mathfrak{b}_{x'} \mid x' > x\}$ for every x in X; that is, for every $\epsilon > 0, \beta_1, \cdots, \beta_k$ in B^* and x in X there is an x' in X such that $x' > x$ and $|b_0(\beta_i) - b_{x'}(\beta_i)| < \epsilon$ for $i = 1, \cdots, k$.*

For each x in X let $E_x = \{\mathfrak{b}_{x'} \mid x' > x\}$; since X is directed, any finite number of the E_x have a point in common, so the sets F_x which are obtained by taking the weak* closure of E_x are weak* closed sets with non-empty finite intersections. Since the sphere $\|\mathfrak{b}_x\| \leq K$ is weak* compact, by the condition for this in terms of closed sets some $\mathfrak{b}_0$ exists in all these F_x; the last clause in the lemma is merely a full statement of the fact that $\mathfrak{b}_0$ is in the weak* closure of every E_x.

2. **The principal theorem.** The terms next defined are the ones used in the statement of the theorem and not merely in its proof.

A set Y is called an Abelian semi-group if there is defined for each pair of elements y, y' in Y a sum $y + y'$ in Y such that $y + y' = y' + y$, and such that $y + (y' + y'') = (y + y') + y''$; that is, addition is commutative and associative. If E is a subset of Y, then $E \dot{-} y$ is the set $\{y' \mid y' + y \in E\}$. It is clear that if $\{E\}$ is any partition of Y into any number of disjoint sets, then $\{E \dot{-} y\}$ is also such a partition.

Let X be a directed set and Y an Abelian semi-group; for each x in X let Ψ_x be a bounded, additive set function over Y. For each y in Y and

E_1 and E_2 are disjoint subsets of Y. If Ψ is bounded $V\Psi(Y) = \sup \sum_{i \leq k} |\Psi(E_i)| \leq 2K$, where the "sup" is taken over all partitions of Y into a finite number of disjoint subsets E_i.

(7) The pertinent properties are these: (1) If $\phi = \sum_{i \leq k} \alpha_i \phi_{E_i}$, $\int \phi d\Psi = \sum_{i \leq k} \alpha_i \Psi(E_i)$. (2) If $\|\phi_n - \phi\|_{M_Y} \to 0$, then $\int \phi_n d\Psi \to \int \phi d\Psi$. See Fichtenholz and Kantorovich [7].

(8) Integrals whose values lie in B^{**} instead of in B were introduced by N. Dunford, *Uniformity in linear spaces*, these Transactions, vol. 44 (1938), pp. 305–356, and I. Gelfand, *Abstrakte Funktionen und lineare Operatoren*, Recueil Mathématique, (n.s.), vol. 4 (1938), pp. 235–284.

x in X let Ψ_{xy} be the bounded additive set function over Y defined by $\Psi_{xy}(E)=\Psi_x(E\dot{-}y)-\Psi_x(E)$ for every $E\subset Y$; the system of set functions $\{\Psi_x\}$ is called nearly invariant over Y if $\lim_x V\Psi_{xy}(Y)=0$ for each y in Y, if $V\Psi_x(Y)$ is bounded, and if $\lim_x \Psi_x(Y)=1$. Ψ is called invariant if $\Psi(E\dot{-}y)=\Psi(E)$ for every $E\subset Y$ and $\Psi(Y)=1$.

If Y is an Abelian semi-group and B is a Banach space and if for each y in Y, T^y is an element of $\mathfrak{T}$, the space of linear operators on B to B, the semi-group $\{T^y\}$ is called a bounded representation of Y if $T^yT^{y'}=T^{y+y'}$ for every y and y' in Y, and if there is a $K\geqq 0$ such that $\|T^y\|\leqq K$ for every y. If $\{T^y\}$ is any bounded representation of Y, let B' and B'' be the subsets of B defined by $B''=\{b \mid T^yb=b$ for every y in $Y\}$ and B' is the smallest closed linear subset of B containing all the points $b-T^yb$ for every b in B and y in Y; M is the smallest linear subset of B containing B' and B''. Clearly, $b\in M$ if and only if one (or all) $T^yb\in M$, since $b=T^yb+(b-T^yb)$ and $b-T^yb$ is in B'.

THEOREM 1. *Let Y be any Abelian semi-group, B a Banach space and $\{T^y\}$ a bounded representation of Y in $\mathfrak{T}$. Let X be a directed set and $\{\Psi_x\}$ a nearly invariant system of set functions over Y, and for each b in B and x in X let $\tau_xb=\int f^b d\Psi_x$ where f^b in $M_Y(B)$ is defined by $f^b(y)=T^yb$ for each y in Y; for each b in B let τb be one of the points which Lemma* 1 *associates with the points τ_xb. Then*:

(1) *τb can be taken in B instead of merely in B^{**} if and only if b is in M.*

(2) *M is a closed linear subset of B and is the direct sum of B' and B'', that is, each b in M is the sum of a b' in B' and a b'' in B'', where b' and b'' are uniquely determined; in fact $b''=\tau b$.*

(3) *τb is uniquely determined if $b\in M$ and, in M, τ is a linear operator with values in B''.*

(4) *If $b\in M$, $\tau\tau b=\tau b$ and $T^y\tau b=\tau b=\tau T^yb$ for every y in Y.*

(5) *$\tau b=b$ if and only if $b\in B''$; $\tau b=\theta$, the zero element in B, if and only if $b\in B'$.*

(6) *If $b\in M$, $\|\tau_xb-\tau b\|_{B^{**}}\to 0$.*

The main body of the proof will be divided into a number of simple steps.

(a) $T^y\tau_xb=\tau_xT^yb$ for every x in X, y in Y and b in B.

For each β in B^*,

$$T^y\tau_xb(\beta) = \tau_xb(T^{y*}\beta) = \int T^{y*}\beta f^b d\Psi_x = \int \beta f^{T^yb} d\Psi_x = \tau_xT^yb(\beta)$$

since

$$T^{y*}\beta f^b(t) = T^{y*}\beta(T^tb) = \beta(T^yT^tb) = \beta(T^tT^yb) = \beta f^{T^yb}(t)$$

for each t in Y.

(b) $\|T^y\tau_xb-\tau_xb\|\to 0$ for each b in B and y in Y(9).

(9) It should be noted that the condition that $V\Psi_{xy}(Y)\to 0$ for each y which is used in the

$$\|T^y\tau_x b - \tau_x b\| = \|\tau_x T^y b - \tau_x b\| = \sup_{\|\beta\|\le 1} |\tau_x T^y b(\beta) - \tau_x b(\beta)|$$

$$= \sup_{\|\beta\|\le 1}\left|\int \beta f^{T^y b} d\Psi_x - \int \beta f^b d\Psi_x\right|$$

$$= \sup_{\|\beta\|\le 1}\left|\int \beta f^b d\Psi_{xy}\right| \le \|f^b\| V\Psi_{xy}(Y)$$

which tends to zero for each y. The only difficulty is in justifying the last equality which can be done as follows: since $\beta f^b \in M_Y$ for each β in B^* and b in B, it suffices to prove that $\int\phi d\Psi_{xy} = \int\phi_y d\Psi_x - \int\phi d\Psi_x$ where ϕ_y is defined by $\phi_y(y') = \phi(y+y')$ and ϕ is any function in M_Y. If ϕ is a simple function; that is, $\phi = \sum_{i\le k}\alpha_i\phi_{E_i}$ where the E_i are disjoint subsets of Y and the α_i are real numbers; then $\int\phi d\Psi_{xy} = \sum_{i\le k}\alpha_i\Psi_{xy}(E_i) = \sum_{i\le k}\alpha_i\Psi_x(E_i \dot- y) - \sum_{i\le k}\alpha_i\Psi_x(E_i) = \int\phi_y d\Psi_x - \int\phi d\Psi_x$ since $\phi_y = \sum_{i\le k}\alpha_i\phi_{E_i\dot- y}$; since the simple functions are dense in M_Y the same is true for any ϕ in M_Y.

(c) $T^y\tau b = \tau b$ for any y in Y and b in B.

It suffices to show for any $\epsilon > 0$ and β in B^* that $|T^y\tau b(\beta) - \tau b(\beta)| < \epsilon$. But

$$|T^y\tau b(\beta) - \tau b(\beta)| \le |T^y\tau b(\beta) - T^y\tau_x b(\beta)|$$
$$+ |T^y\tau_x b(\beta) - \tau_x b(\beta)| + |\tau_x b(\beta) - \tau b(\beta)|;$$

by (b) the middle term is less than $\epsilon/3$ if $x > x_0$. The first term is equal to $|\tau b(T^{y*}\beta) - \tau_x b(T^{y*}\beta)|$; by the definition of τb and Lemma 1, τb is in the weak* closure of $\{\tau_x b \mid x > x_0\}$ so there is an $x > x_0$ such that $|\tau_x b(\beta) - \tau b(\beta)| < \epsilon/3$ and $|\tau_x b(T^{y*}\beta) - \tau b(T^{y*}\beta)| < \epsilon/3$. For this x all three terms are less than $\epsilon/3$ and (c) follows.

(d) If $d \in B'$, then $\|\tau_x d\| \to 0$.

If $d = b - T^y b$ for some b in B and y in Y then $\|\tau_x d\| = \|T^y\tau_x b - \tau_x b\| \to 0$ by (b). Since all the operations in question are additive and homogeneous, $\|\tau_x d\| \to 0$ if $d = \sum_{i\le k}\alpha_i(b_i - T^{y_i}b_i)$ for any choice of y_i in Y, b_i in B, and α_i real. If $d \in B'$ then there is a d_ϵ of this last form such that $\|d - d_\epsilon\| < \epsilon$; then $\|\tau_x d\| < \|\tau_x(d - d_\epsilon)\| + \|\tau_x d_\epsilon\| < (K+1)\epsilon$, where K is the upper bound of $V\Psi_x(Y)$, if x is large enough.

(e) $\|\tau_x\tau b - \tau b\| \to 0$ for each b in M', where M' is the set of those b in B such that τb can be chosen in B, not only in B^{**}.

---

definition of near invariance is stronger than the corresponding condition used by Alaoglu and Birkhoff [3]. Since this property is used only in this step of the proof of the theorem, it is easily seen that the property that must be required of the system $\{\Psi_x\}$ is that $\|\int f^b d\Psi_{xy}\| \to 0$ for each b and y. This shows that the hypotheses on $\{\Psi_x\}$ can be weakened if some restriction is placed on the semi-group $\{T^y\}$; for example, it is sufficient that $\Psi_{xy}(E) \to 0$ for each $E \subset Y$ provided that the semi-group $\{T^y\}$ is so restricted that each f^b is integrable in the sense used in [3]. It seems to me that the extra restrictions are more properly placed on the system $\{\Psi_x\}$ than on the semi-group $\{T^y\}$ since the results are to a large extent independent of all but the existence of $\{\Psi_x\}$ and since (see Theorem 3) a nearly invariant system always exists.

$\tau_x\tau b(\beta)=\int\beta f^{\tau b}d\Psi_x$, but by (c) $f^{\tau b}(y)=T^y\tau b=\tau b$ for each y so $\tau_x\tau b(\beta)=\int\beta(\tau b)\phi_Y d\Psi_x=\Psi_x(Y)\beta(\tau b)=\beta(\Psi_x(Y)\tau b)$, and $\|\tau_x\tau b-\tau b\|=\|\tau b\|\,|\Psi_x(Y)-1|$ which tends to 0 by the hypotheses on $\{\Psi_x\}$.

(f) $b-\tau b\in B'$ for each b in M' so $\|\tau_x b-\tau_x\tau b\|\to 0$ for each b in M'.

If $b_1=b-\tau b$ is in $B-B'$, there exists a β_0 in B^* for which $\beta_0(d)=0$ if $d\in B'$ while $\beta_0(b_1)=1$. Then $\tau_x b(\beta_0)=\int\beta_0 f^b d\Psi_x$, but $\beta_0 f^b(y)=\beta_0(T^y b)=\beta_0(b)$ for each y so $\tau_x b(\beta_0)=\Psi_x(Y)\beta_0(b)$ and $\lim_x \tau_x b(\beta_0)=\lim_x \Psi_x(Y)\beta_0(b)=\beta_0(b)$. On the other hand, for each x in X and $\epsilon>0$ there is an x' in X such that $x'>x$ and $|\tau_{x'}b(\beta_0)-\beta_0(\tau b)|<\epsilon$, so $\beta_0(\tau b)=\beta_0(b)$ and $\beta_0(b_1)=0$ contrary to the choice of β_0.

The conclusions mentioned can now be drawn. (1) is true if $M'=M$. $B'\subset M'$ by (d); if $b\in B''$, $\tau_x b=\Psi_x(Y)b$ so $\|b-\tau_x b\|\to 0$ and $B''\subset M'$; hence $M'\supset M$. If $b\in M'$, by (c) $\tau b\in B''$ and $b-\tau b\in B'$ by (f); hence $b=b+(b-\tau b)\in M$ if $b\in M'$ and $M'\subset M$. (6) follows from (e), (f), and (1). For (3) note that (6) implies that τb is uniquely determined if $b\in M$; τ is clearly additive on M and for each b in M $\|\tau b\|\leqq\sup_x\|\tau_x b\|=\sup_x\|\int f^b d\Psi_x\|\leqq\|f^b\|\sup_x V\Psi_x(Y)$ $\leqq\|b\|\sup_y\|T^y\|\sup_x V\Psi_x(Y)$, so τ is linear on M. All of (2) is proved except that M is closed; let $N=\{b\mid \tau_0 b=\lim_x\tau_x b$ exists in the norm topology in $B^{**}\}$. Then N, as the set of points of convergence of $\{\tau_x\}$, is a closed linear manifold and M is the inverse image by τ_0 of B (considered as imbedded in B^{**}). Hence M is closed in N and therefore in B. For (5) the remarks above showed that $\tau b=b$ if $b\in B''$; if $\tau b=b$, $T^y b=T^y\tau b=\tau b=b$ by (c), so $b\in B''$. (d) shows that $\tau b=\theta$ if $b\in B'$; if $\tau b=\theta$, $b=b-\tau b\in B'$ by (f) since $b\in M$ in this case. For (4) $\tau b\in B''$ if $b\in M$ so $\tau\tau b=\tau b$ and τ is idempotent; $T^y\tau b=\tau b$ by (c); if $b\in M$, $T^y b=(T^y b-b)+b\in M$ too, since $T^y b-b\in B'\subset M$; then $\tau T^y b=\lim_x\tau_x T^y b$ $=\lim_x T^y\tau_x b=T^y(\lim_x\tau_x b)=T^y\tau b=\tau b$.

3. **Related and supplementary theorems.** One point stands out strongly in this set of conclusions: most of them do not directly involve X or $\{\Psi_x\}$ except in so far as the existence of the nearly invariant system $\{\Psi_x\}$ was required to prove existence here. M is already defined in terms of the bounded Abelian semi-group $\{T^y\}$ and (5) defines τ in M in terms of $\{T^y\}$ alone, so that M and τ are the same no matter what X and $\{\Psi_x\}$ are used so long as the system $\{\Psi_x\}$ is nearly invariant over Y. This raises three questions:

(A) If Y is a given Abelian semi-group, is there a nearly invariant system $\{\Psi_x\}$ for some X?

(B) Under what conditions on Y and $\{T^y\}$ does $M=B$?

(C) If some natural choice of X and $\{\Psi_x\}$ is suggested by the nature of Y, is this system nearly invariant?

(A) can be completely answered (yes, for any X); (B) partially; (C) depends on the case in question and obviously has no general answer. This section contains the discussion of (A) and (B).

THEOREM 2. *If a nearly invariant system $\{\Psi_x\}$ of set functions over Y exists*

then there is an invariant function Ψ; that is, a Ψ such that $\Psi(E\dot{-}y)=\Psi(E)$ for every $E\subset Y$ and $y\in Y$ and $\Psi(Y)=1$.

Proof. If $\{\Psi_x\}$ is the nearly invariant system, let Υ_x be the element of M_Y^* defined by $\Upsilon_x(\phi)=\int\phi d\Psi_x$ for every $\phi\in M_Y$. Applying Lemma 1 with $B^*=M_Y$, $\mathfrak{b}_x=\Upsilon_x$ gives an Υ in M_Y^* such that Υ is in the weak* closure of every set $\{\Upsilon_x \mid x>x_0\}$ for every x_0 in X. Let $\Psi(E)=\Upsilon(\phi_E)$ for each $E\subset Y$, then Ψ is additive and bounded, of course, since $V\Psi(Y)\leqq\lim\sup_x V\Psi_x(Y)$ and, since $\lim_x \Psi_x(Y)=1$, $\Psi(Y)=1$.

To show Ψ invariant it suffices to show that for each $\epsilon>0$ $|\Psi(E\dot{-}y)-\Psi(E)|<\epsilon$. Now

$$\begin{aligned}|\Psi(E\dot{-}y)-\Psi(E)| &\leqq |\Psi(E\dot{-}y)-\Psi_x(E\dot{-}y)|+|\Psi_x(E\dot{-}y)-\Psi_x(E)|\\ &\quad+|\Psi_x(E)-\Psi(E)|\\ &=|\Upsilon(\phi_{E\dot{-}y})-\Upsilon_x(\phi_{E\dot{-}y})|+|\Upsilon_x(\phi_{E\dot{-}y})-\Upsilon_x(\phi_E)|\\ &\quad+|\Upsilon_x(\phi_E)-\Upsilon(\phi_E)|.\end{aligned}$$

By the near invariance of $\{\Psi_x\}$, there is an x_0 such that the middle term is less than $\epsilon/3$ whenever $x>x_0$; then by the fact that Υ is in the weak* closure of $\{\Upsilon_x|x>x_0\}$ the other two terms can be made less than $\epsilon/3$ by proper choice of $x>x_0$, so $\Psi(E\dot{-}y)=\Psi(E)$ for all $E\subset Y$ and $y\in Y$.

Note that the full strength of near invariance is not used in this proof but only that $\Psi_{xy}(E)\to 0$ for every $E\subset Y$. Naturally if an invariant function Ψ exists, the system $\{\Psi_x\}$ such that every $\Psi_x=\Psi$ for each x in X, is nearly invariant over Y no matter what X is.

COROLLARY 1. *If Y is a family of subsets y of a given set A, with addition in Y ordinary addition as subsets of A, then there exists an invariant function Ψ over Y.*

This can be proved directly by proper application of the Hahn-Banach theorem. To derive it from Theorem 3, let $X=Y$, order Y by letting $y>y'$ mean $y\supset y'$, and let Ψ_y for each y in Y be defined on the subsets E of Y by $\Psi_y(E)=1$ if $y\in E$, $\Psi_y(E)=0$ if not. To show that the system $\{\Psi_y\}$ is nearly invariant, for any y_0 in Y let $y\supset y_0$; then $\Psi_y(E\dot{-}y_0)=1$ if and only if y is in $E\dot{-}y_0$, that is, if and only if $y+y_0$ is in E, but $y+y_0=y$ if $y\supset y_0$ so $\Psi_y(E\dot{-}y_0)=1$ if and only if $y\in E$, i.e., if and only if $\Psi_y(E)=1$. Therefore $\Psi_y(E)=\Psi_y(E\dot{-}y_0)$ for every E or $V\Psi_{yy_0}(Y)=0$ if $y>y_0$.

This particular nearly invariant system will be used later.

COROLLARY 2. *To the conclusions of Theorem 1 can be added:*

(7) *There exists an invariant function Ψ defined over Y such that $\tau b=\int f^b d\Psi$ if $b\in M$, while $\int f^b d\Psi\in B^{**}$ if $b\notin M$. Moreover $\int f^b d\Psi$ is in the weak* closure of $\{\tau_x b \mid x>x_0\}$ for every x_0 in X.*

Letting $\Psi_x'=\Psi$ for any X gives a nearly invariant system; $\tau' b=\tau_x' b=\int f^b d\Psi$

shows that $\tau' b \in B$ if and only if $b \in M$ (by Theorem 1) and in M $\tau' b = \tau b$. The definition of Υ, and hence of Ψ, in Theorem 2 shows that $\int f^b d\Psi$ lies in the given sets.

The important result is

THEOREM 3. *If Y is an Abelian semi-group, there is an invariant set function Ψ over Y such that $\Psi(Y) = V\Psi(Y) = 1$ and $\Psi(E) \geqq 0$ if $E \subset Y$.*

The proof is essentially that of a result of Morse and Agnew [1, Lemma 2.01]. Let Υ_0 be defined on the multiples of ϕ_Y by $\Upsilon_0(\alpha\phi_Y) = \alpha$; then by the Hahn-Banach theorem there is an Υ_1 in M_Y^* such that $\|\Upsilon_1\| = 1$ and $\Upsilon_1(\alpha\phi_Y) = \Upsilon_0(\alpha\phi_Y)$. For each y in Y let T^y be the function on M_Y to M_Y defined by $T^y\phi(y') = \phi(y+y')$ for each ϕ in M_Y and y' in Y. $\|T^y\| = 1$ for each y since $\sup_{y' \in Y} |\phi(y+y')| \leqq \sup_{y' \in Y} |\phi(y')|$ for every y in Y and $T^y\phi_Y = \phi_Y$; Υ_0 is invariant under all T^y. Following [1] let

$$p(\phi) = \inf \left[\sup_{y_0 \in Y} \frac{1}{n} \sum_{i=1}^{n} \Upsilon_1(T^{y_0} T^{y_i} \phi)\right]$$

where the "inf" is taken over all choices of the integer n and the points $y_1, \cdots, y_n$ in Y. As in [1] it can be shown that $p(\alpha\phi) = \alpha p(\phi)$ if $\alpha \geqq 0$, that $p(\phi + \phi') \leqq p(\phi) + p(\phi')$, and that $p(\phi) \leqq \|\phi\|$ for every ϕ in M_Y. Since $\Upsilon_0(\alpha\phi_Y) \leqq p(\alpha\phi_Y)$, applying the Hahn-Banach theorem again gives an Υ in M_Y^* such that $\Upsilon(\alpha\phi_Y) = \alpha$ and $\Upsilon(\phi) \leqq p(\phi) \leqq \|\phi\|$ for every ϕ; the proof that Υ is invariant under the T^y is as in [1].

Let $\Psi(E) = \Upsilon(\phi_E)$ if $E \subset Y$. Then $\Psi(E \dot{-} y) = \Upsilon(\phi_{E \dot{-} y}) = \Upsilon(T^y\phi_E) = \Upsilon(\phi_E) = \Psi(E)$, so Ψ is invariant; $\|\Upsilon\| = \Upsilon(\phi_Y) = 1$ so $V\Psi(Y) = \Psi(Y) = 1$. Ψ is non-negative, for if E exists with $\Psi(E) < 0$, then $V\Psi(Y) \geqq |\Psi(E)| + |\Psi(Y-E)| = -\Psi(E) + 1 - \Psi(E) > 1$.

COROLLARY 3. $\|\tau b\|_B \leqq \|f^b\|_{M_Y(B)} = \sup_{y \in Y} \|T^y b\|_B \leqq \|b\| \sup_{y \in Y} \|T^y\|$ *if* $b \in M$.

If $b \in M$, $\tau b = \int f^b d\Psi$ where Ψ is defined by Theorem 3; hence $\|\tau b\| = \|\int f^b d\Psi\| \leqq \|f^b\| V\Psi(Y) = \|f^b\|$.

COROLLARY 4. *If $b \in M$, then τb is in the closed convex hull $K(b)$ of the set $\{T^y b \mid y \in Y\}$.*

For each β in B^*, $\beta(\tau b) = \int \beta f^b d\Psi \leqq V\Psi(Y) \sup_{y \in Y} \beta f^b(y) = \sup_{y \in Y} \beta(T^y b)$. If b_0 is not in $K(b)$ there is a β_0 such that $\beta_0(b_0) > \sup_y \beta_0(T^y b)$ (by a theorem of Mazur [9]) so τb is in $K(b)$.

Alaoglu and Birkhoff call a point b in B ergodic relative to the bounded semi-group $\{T^y\}$ if there is just one point of B'' in $K(b)$ and all $K(T^y b)$(10).

(10) This is not their definition but one of the properties shown to be necessary and sufficient.

COROLLARY 5. *b is ergodic if and only if $b\in M$.*

If $b\in M$, $\tau b\in B''$ and to $K(b)$ by Theorem 2 and Corollary 4; all $T^y b$ are in M too and $\tau T^y b=\tau b$, so τb is all $K(T^y b)$ also and b is ergodic if these sets contain no other points in B''; this is asserted by Lemmas 1 and 2 of [2]. If b is ergodic and b_0 is the unique point in $B''\cdot K(b)$, then θ is the unique point in $B''\cdot K(b-b_0)$; let $b_1=b-b_0$. Since θ is in the closed convex hull of the set $\{T^y b_1 \mid y\in Y\}$, for each $\epsilon>0$ there is a point $b_2=\sum_{i\leq k}\alpha_i T^{y_i} b_1$ with $\|b_2\|<\epsilon/K$, where $K=\sup_{y\in Y}\|T^y\|$. If $\tau b'=\int f^{b'} d\Psi$, with Ψ as in Theorem 3, then $\tau b_2=\sum_{i\leq k}\alpha_i \tau T^{y_i} b_1=\tau b_1$ and $\|\tau b_1\|=\|\tau b_2\|<\epsilon$. Hence $\|\tau b_1\|=0$ or $b_1\in B'$; so $b=b_1+b_0\in M$.

To complete the relation of these results to those of Alaoglu and Birkhoff requires some study of Problem (B).

LEMMA 3. *If $f\in M_Y(B)$ and Ψ is a bounded additive set function over Y, $\int fd\Psi$ is in Y if P, the smallest closed linear subspace of B containing all the points $\{f(y) \mid y\in Y\}$, is reflexive.*

Each β in B^* defines a π_β in P^* by $\pi_\beta(p)=\beta(p)$ for each p in P; then $\int fd\Psi(\beta)=\int \beta fd\Psi=\int \pi_\beta fd\Psi$. Since the π_β cover P^* this defines a unique p in P^{**} such that $p(\pi)=\int \pi fd\Psi$ if $\pi\in P^*$. Since P is reflexive there is a p in P such that $\pi(p)=\int\pi fd\Psi$ if $\pi\in P^*$. Hence $\beta(p)=\pi_\beta(p)=\int\pi_\beta fd\Psi=\int\beta fd\Psi$; so $\int fd\Psi=p\in P\subset B$.

COROLLARY 6. *If the set $\{T^y b \mid y\in Y\}$ lies in some reflexive subspace of B, then $b\in M$.*

COROLLARY 7. *If B is reflexive, $M=B$.*

COROLLARY 8. *If B is reflexive and $\{T^y\}$ is any bounded Abelian semi-group of operators on B to B, then each b in B is ergodic; that is, for each b in B $K(b)$ contains just one fixed point τb of all the T^y.*

It is to be noticed that this is a great strengthening of Theorem 5 of [2] since, as is known, every uniformly convex space is reflexive, and since $\|T^y\|\leq 1$ can be replaced by $\|T^y\|\leq K$. On the other hand, that result can be proved with far less machinery.

Corollaries 3 and 7 together imply that if B is reflexive the set B'' of common fixed points of any bounded Abelian semi-group $\{T\}$ of operators on B to B is the range of a projection operator τ defined on all of B and $\|\tau\|\leq\sup\|T\|$.

A trivial result is this:

If $\{T^y\}$ is any bounded Abelian semi-group of operators on B to B and if there is a y_0 such that $\|T^{y_0}\|<1$, then $B'=B$; in fact $\lim_y\|T^y\|=0$ if y is directed by letting $y>y'$ if there is a y'' such that $y=y''+y'$.

For each $\epsilon>0$, there is an n such that $\|(T^{y_0})^n\|<\epsilon/K$, where K is the bound of the norms of all T^y, then $\|T^y(T^{y_0})^n\|<\epsilon$ for every y or $\|T^y\|<\epsilon$ if $y>ny_0$.

In some cases a certain X and a nearly invariant system $\{\Psi_x\}$ over Y arise naturally. Under certain conditions on X the reflexivity condition of Corollary 7 can be weakened (at least formally). A directed set X has a *countable cofinal subset* if there is a countable subset X' of X such that each x in X is followed by some x' in X'.

THEOREM 4. *If $\{T^y\}$ is a bounded representation of the Abelian semi-group Y, if X is a directed set with a countable cofinal subset, and if the system $\{\Psi_x\}$ is nearly invariant over Y, then $b \in M$ if and only if there is a countable cofinal subset $\{x_n\}$ of X and a b_0 in B such that $\lim_n \tau_{x_n} b(\beta) = \beta(b_0)$ for every β in B^*. In this case $b_0 = \tau b$.*

For such an X this sequential compactness condition assures that τb is in B; that is, that b is in M. Since norm convergence implies weak convergence this condition is satisfied if $b \in M$.

COROLLARY 9. *If B is a Banach space with sequentially weakly compact*(11) *unit sphere, if $\{T^y\}$ is a bounded Abelian semi-group of operators on B to B, if X has a countable cofinal subset and if $\{\Psi_x\}$ is a nearly invariant system such that $\int f^b d\Psi_x \in B$ for each b in B, then $M = B$.*

Since every reflexive space has a sequentially weakly compact unit sphere, this result is related to Corollary 7; since it is not known whether or not sequential weak compactness implies reflexivity, it is not known whether the hypotheses on X and $\{\Psi_x\}$ are needed.

4. **Special semi-groups and systems of set functions.** A theorem of Dunford [6] uses E_n, euclidean n-space with coordinates $y_1, \cdots, y_n$, for Y and the class of n-dimensional cubes $x = \{y \mid \alpha_j < y_j < \alpha_j + r, j = 1, \cdots, n\}$, where $r > 0$ and the α_j are arbitrary real numbers, for X, defining $\Psi_x(E) = m(Ex)/m(x)$, where m is Lebesgue measure, for every Lebesgue measurable set $E \subset Y$. (He has then to assume measurability for each f^b in order to integrate.) $\tau_x b$, then, is the arithmetic mean of f^b over the cube x; that is, $\tau_x b = r^{-n} \int_x f^b \, dm$, where r is the length of edge of the cube x. X is ordered by the size of the cubes, that is, $x > x'$ if the edges of x are longer than those of x'.

A more general result follows from a simple property of convex bodies with interior points in E_n. In what follows let $S_\alpha(y)$ be the closed sphere about y of radius α: as in any linear space if $E, E' \subset E_n$ let $E + E' = \{e + e' \mid e \text{ in } E \text{ and } e' \text{ in } E'\}$ and for any real α and $E \subset E_n$ let $\alpha E = \{\alpha e \mid e \text{ in } E\}$.

LEMMA 4. *If E is a convex subset of E_n and if E contains a sphere $S_r = S_r(0)$, and if $S_\alpha = S_\alpha(0)$, then $S_\alpha + E \subset [(r+\alpha)/r]E$*(12).

---

(11) A set $B_0 \subset B$ is sequentially weakly compact if for each sequence $\{b_n\} \subset B_0$ there is a subsequence $\{n_i\}$ and a b_0 in B_0 such that $\lim_i \beta(b_{n_i}) = \beta(b_0)$ for every β in B^*.

(12) This proof is due to the referee who notes that it holds in any normed vector space.

$y \in S_\alpha$ if and only if $y = (\alpha/r)y'$ for some y' in S_r. For each y in S_α and y'' in E

$$y + y'' = \frac{\alpha}{r} y' + y'' = \frac{r+\alpha}{r}\left[\frac{\alpha}{r+\alpha} y' + \frac{r}{r+\alpha} y''\right] = \frac{r+\alpha}{r} y_0$$

where $y_0 \in E$ since $y', y'' \in E$, $\alpha/(r+\alpha) + r/(r+\alpha) = 1$, and $\alpha/(r+\alpha) > 0$, $r/(r+\alpha) > 0$. Hence $E \dotplus S_\alpha \subset [(r+\alpha)/r]E$.

For each bounded convex set E with interior points contained in E_n let $r(E)$ be the least upper bound of the radii of the spheres contained in E. Then there will be at least one sphere of radius $r(E)$ contained in $\overline{E}$, the closure of E, since any bounded closed set in E_n is compact.

LEMMA 5. *If X is the set of all bounded convex sets with interior points in E_n and if X is directed by the relation $x > x'$ if and only if $r(x) \geqq r(x')$, then for each y in E_n, $\lim_x m[x(x \dot{-} y)]/m(x) = 1$.*

Since $x \dot{-} y = x(x \dot{-} y) + [(x \dot{-} y) - x]$, it suffices to show that $m[(x \dot{-} y) - x]/m(x) \to 0$. If α is the distance from y to 0, then $x \dot{-} y \subset x \dotplus S_\alpha$; by Lemma 4, $x \dotplus S_\alpha$ is contained in a dilation of $\bar{x}$ in the ratio $(r(x)+\alpha)/r(x)$ about the center of any sphere of radius $r(x)$ contained in x. The ratio of the measures of the dilated set and x is precisely $[(r(x)+\alpha)/r(x)]^n$ so $m[(x \dot{-} y) - x]/m(x) \leqq m[(x \dotplus S_\alpha) - x]/m(x) \leqq [(r(x)+\alpha)/r(x)]^n - 1$ which tends to 0 as $r(x) \to \infty$.

LEMMA 6. *There is a non-negative additive set function μ defined on all subsets of E_n of finite outer measure, such that $\mu(\alpha A \dot{-} y) = |\alpha| \mu(A)$ and such that $\mu(A) = m(A)$ if A is Lebesgue measurable and of finite measure.*

This follows from the work of Morse and Agnew [1]; they gave the construction for the case $n = 1$.

THEOREM 5. *Let A be any convex set with interior points in E_n and let $Y = \{\alpha y \mid y \text{ in } A \text{ and } \alpha \geqq 1\}$; then Y is an Abelian semi-group under vector addition in E_n; for any Banach space B let $\{T^y\}$ be a bounded representation of Y. Let X be the set of all convex sets of finite nonzero measure contained in Y and for each x in X define Ψ_x by $\Psi_x(E) = \mu(Ex)/\mu(x)$. Then the system $\{\Psi_x\}$ is nearly invariant and the conclusions of Theorem 1 hold.*

Near invariance of $\{\Psi_x\}$ is all that needs be verified. $V\Psi_x(Y) = \Psi_x(Y) = \mu(x)/\mu(x) = 1$;

$$\begin{aligned} \Psi_{xy}(E) &= \Psi_x(E \dot{-} y) - \Psi_x(E) = [\mu(x)]^{-1}\{\mu[(E \dot{-} y)x] - \mu(Ex)\} \\ &= [\mu(x)]^{-1}\{\mu[(E \dot{-} y)(x \dot{-} y)] - \mu[(E \dot{-} y)((x \dot{-} y) - x)] \\ &\quad + \mu[(E \dot{-} y)(x - (x \dot{-} y))] - \mu(Ex)\} \\ &= [\mu(x)]^{-1}\{\mu(Ex \dot{-} y) - \mu(Ex) - \mu[(E \dot{-} y)((x \dot{-} y) - x)] \\ &\quad + \mu[(E \dot{-} y)(x - (x \dot{-} y))]\} \\ &= [\mu(x)]^{-1}\{-\mu[(E \dot{-} y)((x \dot{-} y) - x)] + \mu[(E \dot{-} y)(x - (x \dot{-} y))]\}. \end{aligned}$$

If $Y=\sum_{i\leq k}E_i$ with the E_i disjoint, then the sets $E_i\dot{-}y$ are disjoint and have sum Y. Hence

$$\begin{aligned}\sum_{i\leq k}|\Psi_{xy}(E_i)| &\leq [\mu(x)]^{-1}\sum_{i\leq k}\{\mu[(E_i\dot{-}y)((x\dot{-}y)-x)] \\ &\qquad + \mu[(E_i\dot{-}y)(x-(x\dot{-}y))]\} \\ &\leq [\mu(x)]^{-1}\left\{\mu\left[\sum_{i\leq k}(E_i\dot{-}y)((x\dot{-}y)-x)\right]\right. \\ &\qquad \left. + \mu\left[\sum_{i\leq k}(E_i\dot{-}y)(x-(x\dot{-}y))\right]\right\} \\ &\leq [\mu(x)]^{-1}\{\mu[(x\dot{-}y)-x]+\mu[x-(x\dot{-}y)]\}.\end{aligned}$$

By Lemma 3, this is small for $r(x)$ large, independent of the decomposition of Y into the sets E_i, so $V\Psi_{xy}(Y)\rightarrow 0$ and this system $\{\Psi_x\}$ is nearly invariant.

The set X' of cubes used by Dunford if ordered by edge length has the same ordering as if ordered by the radius of the largest sphere inside; so X' is a cofinal subset of this family X of convex sets of finite, nonzero measure; so $\lim_{x'}\tau_{x'}b$ exists for every b in M since $\lim_x \tau_x b$ exists for such b; moreover if the functions f^b are all measurable, the $\tau_{x'}$ reduce to Dunford's transformations and Theorem 5 offers a simple proof, without differentiation theorems, of Dunford's result.

Note that this X has a countable cofinal subset, in fact any sequence $\{x_n\}\subset X$ such that $r(x_n)\rightarrow\infty$ will do. Hence if the assumption is made that each $T^y b$ is a measurable function, each $\tau_x b\in B$ and Corollary 9 can be applied with proper choice of B.

For a second application (not considered anywhere in the literature so far as I know) take Y to be the *stack* Δ whose elements are the finite subsets of some given set D of elements d where addition is, as in Corollary 1, ordinary point set addition. If B is a Banach space and $\{T^\delta \mid \delta\in\Delta\}$ is a bounded representation of Δ in the space of linear operators on B to B, $T^\delta T^\delta = T^{\delta+\delta}=T^\delta$ so every T^δ must be a projection; moreover $T^\delta=\prod_{d\in\delta}T^d$ for each $\delta\in\Delta$. Δ is also a directed set if $\delta>\delta'$ means $\delta\supset\delta'$; for each $\delta\in\Delta$ let Ψ_δ be defined over the subsets of Δ by $\Psi_\delta(E)=1$ if $\delta\in E$, $\Psi_\delta(E)=0$ if $\delta\notin E$. Then, as in Corollary 1, the system $\{\Psi_\delta\}$ is nearly invariant over Δ.

THEOREM 6. *If B is a Banach space and the T^d are commuting projections on B to B such that $\|T^\delta\|\leq K$ for all δ, where $T^\delta=\prod_{d\in\delta}T^d$, then* $\lim_\delta T^\delta b$ *exists (in the norm topology) if and only if $b\in M$ (where M is defined as in Theorem* 1) *and for such b,* $\lim_\delta \|T^\delta b-\tau b\|=0$.

By Theorem 1, $\|\tau_\delta b-\tau b\|\rightarrow 0$ with τb in B if and only if $b\in M$. But $\tau_\delta b=\int f^b d\Psi_\delta = T^\delta b$, and if the $T^\delta b$ converge at all their limit must be in B.

It is to be noted that this theorem has a much stronger conclusion than can be obtained in general; this is true because the bounded representations of a stack are so greatly restricted.

COROLLARY 10. *If B is reflexive and if T^d are commuting projections with $\|T^\delta\|$ uniformly bounded, then* $\lim_\delta T^\delta b = \tau b$ *for every b in B, where τ is a projection defined on all of B which has all the properties ascribed to τ in Theorem* 1.

One more consequence of Theorem 1 is this: If the collection $\{T^y\}$ of operators on B to B is a bounded representation of the semi-group Y, then the collection $\{T^{y*}\}$ is such a representation in the set of operators on B^* to B^*. In general, defining $B^{(n*)}$ by induction from $B^{(n*)} = (B^{((n-1)*)})^*$, and $T^{(n*)} = (T^{((n-1)*)})^*$, the same is true of the collection $\{T^{y(n*)}\}$. Hence Theorem 1 defines in $B^{(n*)}$ a set $M^{(n*)}$ consisting of the direct sum of $B''^{(n*)}$, the set of fixed points of the $T^{y(n*)}$, and $B'^{(n*)}$, the smallest closed linear set in $B^{(n*)}$ containing all $T^{y(n*)}b^{(n*)} - b^{(n*)}$ for all choices of y in Y and $b^{(n*)}$ in $B^{(n*)}$. Since $B^{((n+2)*)} \supset B^{(n*)}$ for every n, and the $T^{y((n+2)*)}$ agree with the $T^{y(n*)}$ in $B^{(n*)}$, $B''^{((n+2)*)} \cdot B^{(n*)} = B''^{(n*)}$ and similar relations hold for $B'^{(n*)}$ and $M^{(n*)}$. (c) of the proof of Theorem 1 shows that for any b in B and any possible choice of τb, τb is in $B''^{(2*)}$.

An example of the results obtained from this point of view is

THEOREM 7. *If B is reflexive and if $\{T^y \mid y \in Y\}$ is a bounded Abelian semi-group of operators on B to B, then there exists a projection τ defined over all B such that for every nearly invariant system of set functions $\{\Psi_x\}$ over Y, $\lim_x \|\tau_x b - \tau b\| = 0$ for each b in B and $\lim_x \|\tau_x^* \beta - \tau^* \beta\| = 0$ for each β in B^*. τ and τ^* have in their respective spaces the properties specified by Theorem* 1.

All that need be verified is that τ_0^*, the projection in B^* that exists by direct application of Theorem 1 to that space, is equal to τ^*, the adjoint of τ. Since $\|\tau_x^*\beta - \tau_0^*\beta\| \to 0$ for each β in B^*, $\tau_x^*\beta(b) - \tau_0^*\beta(b) = \beta(\tau_x b) - \tau_0^*\beta(b) \to 0$ for each b in B but $\beta(\tau_x b) \to \beta(\tau b)$ so $\tau^*\beta(b) = \beta(\tau b) = \tau_0^*\beta(b)$ for each b in B and β in B^*, or $\tau^* = \tau_0^*$.

I see no way of proving anything quite of this nature if $M \neq B$.

It may be noted that neither the methods nor the results of this paper carry over to noncommutative semi-groups; in fact, an example [2, §13, Example 1] shows that for non-abelian semi-groups the set M and the set of ergodic points of B need not be the same.

BIBLIOGRAPHY

1. R. P. Agnew and A. P. Morse, *Extensions of linear functionals, with applications to limits, integrals, measures, and densities*, Annals of Mathematics, (2), vol. 39 (1938), pp. 20–30.

2. L. Alaoglu, *Weak topologies in normed linear spaces*, Annals of Mathematics, (2), vol. 41 (1940), pp. 252–267.

3. L. Alaoglu and G. Birkhoff, *General ergodic theorems*, Annals of Mathematics, (2), vol. 41 (1940), pp. 293–309.

4. S. Banach, *Théorie des Opérations Linéaires*, Monografje Matematiczne, Warsaw, 1932.

5. G. Birkhoff, *Moore-Smith convergence in general topology*, Annals of Mathematics, (2), vol. 38 (1937), pp. 39–56.

6. L. W. Cohen, *On the mean ergodic theorem*, Annals of Mathematics, (2), vol. 41 (1940), pp. 505–509.

7. N. Dunford, *A mean ergodic theorem*, Duke Mathematical Journal, vol. 5 (1939), pp. 635–646.

8. G. Fichtenholz and L. Kantorovich, *Sur les opérations dans l'espace des functions bornées*, Studia Mathematica, vol. 5 (1934), pp. 69–98.

9. S. Kakutani, *Weak topology and regularity of Banach spaces*, Proceedings of the Imperial Academy, Tokyo, vol. 16 (1939), pp. 169–173.

10. E. R. Lorch, *Means of iterated transformations in reflexive Banach spaces*, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 945–947.

11. S. Mazur, *Über konvexe Mengen in linearen normierten Räumen*, Studia Mathematica, vol. 4 (1933), pp. 70–84.

12. F. Riesz, *Some mean ergodic theorems*, Journal of the London Mathematical Society, vol. 13 (1938), pp. 274–278.

13. V. Šmulian, *Sur les topologies différents dans l'espace de Banach*, Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS, vol. 23 (1939), pp. 331–334.

14. J. V. Wehausen, *Transformations in linear topological spaces*, Duke Mathematical Journal, vol. 4 (1938), pp. 157–168.

15. K. Yosida, *Mean ergodic theorems in Banach spaces*, Proceedings of the Imperial Academy, Tokyo, vol. 14 (1939), pp. 292–294.

University of Illinois,
Urbana, Ill.

ON MEASURE IN ABSTRACT SETS

BY

DOROTHY MAHARAM

INTRODUCTION

In their paper(1) *Sur les décompositions dénombrables* Banach and Tarski obtained a result which can be restated as follows:

A necessary and sufficient condition that two Lebesgue measurable, euclidean sets a and a', shall have equal measure is that there exist null sets $n<a$ and $n'<a'$, such that $a-n$ and $a'-n'$ are the respective unions of sequences of disjoint measurable sets, of which corresponding sets are congruent.

If we then identify sets differing on only a null set, there exists a class of transformations on the measurable sets such that two sets are of equal measure if and only if they correspond under some transformation of that class. Then just as the elementary notion of the volume of an n-dimensional interval is generalized to that of any measure function on a Borel field, so the equally elementary notion of equality of volume, defined for these figures by the relation congruence, can be generalized to the notion of equality of measure defined for some field of sets by a suitable class of transformations. This is done as follows:

In Part I, we consider a complemented, distributive σ-lattice M with a zero element, and a class Φ of σ-isomorphisms on the principal ideals of M. The lattice M is to be taken to correspond to a family of measurable sets modulo the null sets, and Φ as a semi-group of measure-preserving transformations. Then Φ generates an equivalence relation $a\sim b$ between elements of M which is countably additive and hereditary in the sense that $a\sim b$, $a'<a$ imply that there exists a $b'<b$ such that $a'\sim b'$. It is then shown for the bounded elements of M, that is, those which are not equivalent to any proper subelement, that the relation $a\sim b$ is also preserved by subtraction and by taking limits of monotonic sequences. It may be remarked that these latter results yield an independent proof(2) of the theorem of Banach and Tarski. A construction is also given leading to a definition of a complete measure. In Part II, the measure of an element of M is defined as the totality of its equiva-

Presented to the Society February 22, 1941; received by the editors November 19, 1940, and in revised form, March 15, 1941.

(1) Fundamenta Mathematicae, vol. 6 (1924), pp. 244–277. The theorem quoted is on page 277. This result has been extended to the case of the Haar measure on a locally compact group by Professor John von Neumann in a series of lectures at the Institute for Advanced Study.

(2) See Example 15.3 for a proof in a generalized form.

lent elements, that is, of images under Φ; and a partial ordering and an operation of addition are defined in the set of measures. A necessary and sufficient condition is then given that the set of measure values be order and addition isomorphic to a set of positive numbers. Such an isomorphism is necessarily unique, up to a multiplicative constant. These results will be analyzed in detail in another paper. In Part III, a sufficient condition is given that, for a family of sets upon which a numerical measure is defined, two sets of equal measure shall correspond under some measure-preserving transformation, and so that the given measure shall be derivable by the procedure of Part II. It turns out that this is the case for a measure on a separable Borel field, provided there are no minimum sets of positive measure. However, a trivial generalization of the theory of Part II can be seen to be applicable to any measure function.

PART I

1. **Ideals and isomorphisms in the lattice** M. Let $M \equiv \{a, b, c, \cdots\}$ be a complemented distributive σ-lattice with a zero element; i.e., satisfying the system (α):

i. *M has a partial ordering $a<b$ (b contains a) such that $a<b$ excludes $a=b$.*

ii. *Every sequence $(a_n) \subseteq M$ has a supremum and an infimum $\sum a_n$ and $\prod a_n$, respectively. If (a_n) is monotone increasing (decreasing), we denote $\sum a_n$ ($\prod a_n$)* by $\lim a_n$. *We also write $a+b$ and ab for the supremum and infimum of (a, b).*

iii. *M has a zero element θ. Any elements a, b such that $ab=\theta$ are called disjoint.*

iv. *If $a<b$, then there exists a unique element, which will be written $b-a$, such that $(b-a)a=\theta$, and $a+(b-a)=b$.*

v. *For any sequence (a_n) and any a, $a(\sum a_n)=\sum aa_n$.*

1.1. DEFINITION. *A class $M' \subseteq M$ is called a σ-ideal if $a'<a \in M'$ implies that $a' \in M'$, and if $(a_n) \subset M'$ implies that $\sum a_n \in M'$. The ideal M' is called principal if there exists an element a such that $a' \in M'$ if and only if $a' \leqq a$. The principal ideal corresponding to an element a will be written $I(a)$.*

1.2. DEFINITION. *A univocal correspondence ϕ defined on a σ-ideal $M' \subseteq M$ onto a σ-ideal $M'' \subseteq M$ is called a σ-isomorphism if for every sequence $(a_n) \subset M$, $\phi(\sum a_n) = \sum \phi(a_n)$, and $\phi(\prod a_n) = \prod \phi(a_n)$. The correspondence inverse to ϕ is also a σ-isomorphism which will be written ϕ^{-1}.*

1.21. For any σ-isomorphism ϕ, $\phi(a)=\theta \equiv a=\theta$; $a'<a \equiv \phi(a')<\phi(a)$, and $\phi(a-a')=\phi(a)-\phi(a')$.

1.22. Let $\boldsymbol{m}$ be a σ-ring of subsets of a set $\boldsymbol{s}$; $\boldsymbol{n}$, a σ-ideal in $\boldsymbol{m}$; and M, the residue class ring $\boldsymbol{m}/\boldsymbol{n}$. Then any univocal point correspondence between the subsets $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\boldsymbol{m}$ which leaves the rings $\boldsymbol{m}$ and $\boldsymbol{n}$ invariant leads to a σ-isomorphism of the principal ideals in M corresponding to $\boldsymbol{a}$ and $\boldsymbol{b}$.

1.23. Definition. *Let* (ϕ_n) *be a sequence of* σ*-isomorphisms between the principal ideals* $I(a_n)$ *and* $I(b_n)$. *If the sequences* (a_n) *and* (b_n) *are each disjoint, then* (ϕ_n) *is said to have a sum* $\sum\phi_n$; *namely, the isomorphism* $\phi(a')$ *defined for* $a' \leqq \sum a_n$ *by* $\phi(a') = \sum\phi_n(a'a_n)$.

2. **The class Φ of σ-isomorphisms. Equivalence. Boundedness.** Let Φ be a class of σ-somorphisms each defined between principal ideals of M, and satisfying (β):

i. *The inverse of any isomorphism in* Φ *is in* Φ.

ii. *If* ϕ, ϕ' *are in* Φ, *and are defined on* $I(a)$, $I(a')$, *respectively, then the resultants* $\phi\phi'$ *and* $\phi'\phi$, *defined on* $a\phi'(a')$ *and* $a'\phi(a)$, *are in* Φ.

iii. *If* $(\phi_n) \subset \Phi$, *and if* $\sum\Phi_n$ *exists, then* $\sum\phi_n \in \Phi$.

iv. *At least one isomorphism of* Φ *is defined on any principal ideal of* M.

2.1. Definition. *Two elements of* M *are called* Φ*-equivalent if they correspond under some isomorphism of* Φ. *This relation will be written* $a \sim b$.

2.2. The relation Φ-equivalence has the following properties:

i. *It is an equivalence relation*, and so partitions M into a set of exhaustive and mutually exclusive classes M_α. The unique class containing an element a may be denoted by $M(a)$. Obviously, $M(a) = M(a')$ if and only if there is an isomorphism ϕ of Φ such that $a' = \phi(a)$.

ii. If (a_n), (b_n) are sequences of disjoint elements of M such that $a_n \sim b_n$, then $\sum a_n \sim \sum b_n$; i.e., *Φ-equivalence is countably additive.*

iii. *It is hereditary in the sense that if* $a \sim b$, *and* $a' < a$, *then there exists a* $b' < b$ *such that* $a' \sim b'$.

2.3. Definition. *An element* a *of* M *is unbounded (with respect to* Φ) *if there exists an element* $a' < a$ *such that* $a \sim a'$, *and* $a - a' \neq \theta$. *Otherwise* a *is bounded.*

2.31. i. If a is bounded and $a' < a$, then a' is bounded. ii. If $a \sim a + b$, and $b - ab \neq \theta$, then a is unbounded. iii. If $a \sim a'$, then a and a' are bounded or unbounded together.

2.32. Definition. *A class* M_α *of equivalent elements of* M *is bounded or unbounded according as its elements are either all bounded or all unbounded.*

2.33. Theorem. *A necessary and sufficient condition that an element* a *be unbounded is that* a *contain a sequence of disjoint equivalent elements* $a_n \neq \theta$.

Proof. Suppose such a sequence exists. Then by 2.2, ii, $\sum a_n \sim \sum a_{2n}$; and since $\sum a_{2n-1} \neq \theta$, $\sum a_n$ is unbounded. Conversely, if a is unbounded, there exists an element $a' < a$ and an isomorphism ϕ of Φ such that a' and $a - a'$ are different from θ and $\phi(a) = a - a'$. Then an infinite sequence of elements satisfying the required conditions can be defined inductively by the formula $a_1 = a$, and for $n > 1$, $a_n = \phi(a_{n-1})$.

3. **Equivalence of complements of bounded equivalent elements.** The results of this section can be summarized as follows:

If $a \sim b$, and if either (i) *there exists a σ-isomorphism ϕ of Φ under which ab is invariant, or* (ii) *the elements a and b are bounded, then for any $c>(a+b)$, $c-a\sim c-b$.*

The assertions leading to this statement are the following:

3.1. A necessary and sufficient condition that for any element $c \geqq (a+b)$, $c-a\sim c-b$, is that $a-ab\sim b-ab$. The necessity is obvious, and the sufficiency follows from the fact that $c-(a+b)\sim c-(a+b)$ (identity), and $a-ab\sim b-ab$ (by hypothesis). Since the terms on either side are disjoint, their respective sums are equivalent and $c-a\sim c-b$.

3.2. THEOREM. *If $a\sim b$, and a and b are bounded, then $a-ab\sim b-ab$.*

The general argument of the proof is as follows: First it is shown that there exist sequences (a_n) and (b_n) of elements of M contained in $a-ab$ and $b-ab$, respectively, such that $\phi^n(a_n)<ab$, and $\phi^{n+1}(a_n)=b_n$. From this it is verified that both families are disjoint. Finally it is shown that $\sum a_n = a-ab$ and $\sum b_n = b-ab$, which proves the theorem.

Let us first restate this in the more convenient form:

If $a+x\sim b+x$, where $a+x$, $b+x$ are bounded, and $ax=bx=\theta$, then $a\sim b$.

Proof. We have given that there exists an isomorphism ϕ of Φ such that $\phi(a+x)=b+x$. We shall write this as

(1) $$a + x \sim b + x\ (\phi).$$

Then

(2) $$a\phi^{-1}(b) \sim b\phi\,(a).$$

By taking differences on both sides we get

(3) $$a\phi^{-1}(x) + x \sim b\phi\,(x) + x.$$

If we can show that

(4) $$a\phi^{-1}(x) \sim b\phi\,(x),$$

then by summing corresponding sides of (2) and (4), we get the required result. Now let

(5, i) $$a_0 \equiv a\phi^{-1}(x) \equiv a - a\phi^{-1}(b),$$

(5, ii) $$b_0 \equiv b\phi(x) \equiv b - b\phi(a),$$

(5, iii) $$x_0 \equiv \phi(a_0) \leqq x.$$

Since ϕ is defined on $x \leqq a+b$, we can define inductively for $n>0$

(6) $$x_n \equiv x\phi(x_{n-1}).$$

It can be shown by induction that the elements a, x_0, x_1, $\cdots$, are dis

joint. For since $a_0 \leqq a$, and $x_0 \leqq x$, a_0 and x_0 are disjoint. Suppose that $a_0, x_0, \cdots, x_{n-1}$ are disjoint. But for $0<k\leqq n$, $\phi^{-1}(x_k)\leqq x_{k-1}$, and $\phi^{-1}(x_0)=a_0$. Hence for $0\leqq k\leqq n$, the elements $\phi^{-1}(x_k)$, and so also the elements x_k, are disjoint by the hypothesis of the induction, which is now complete. We define

$$b_n = b\phi(x_{n-1}), \qquad n = 1, 2, 3, \cdots. \tag{7}$$

Since $\phi^{-1}(b_n)\leqq x_{n-1}$, it follows that the elements b_n are disjoint. Further, since $\phi(x_n)\leqq\phi(x)\leqq b+x$,

$$\phi(x_{n-1}) = x\phi(x_{n-1}) + b\phi(x_{n-1}) = x_n + b_n, \tag{8}$$

that is,

$$x_{n-1} \sim x_n + b_n\ (\phi) \qquad \text{for } n = 1, 2, 3, \cdots, \tag{9}$$

and we also have

$$a_0 \sim x_0\ (\phi). \tag{10}$$

Since the terms on the right (the left) side of (9) and (10) are disjoint, their respective sums are equivalent and

$$\left(a_0 + \sum_0^\infty x_n\right) \sim \left(\sum_0^\infty x_n + \sum_1^\infty b_n\right)(\phi). \tag{11}$$

But from (3) and (5)

$$a_0 + x \sim b_0 + x\ (\phi), \tag{12}$$

and for $n\geqq 1$, $b_n\leqq b\phi(x)=b_0$, whence the respective differences of corresponding sides of (11) and (12) are equivalent; that is,

$$x - \sum_0^\infty x_n \sim \left(b_0 - \sum_1^\infty b_n\right) + \left(x - \sum_0^\infty x_n\right). \tag{13}$$

Since $a+x$ is bounded, so is $x-\sum_0^\infty x_n$, and hence

$$b_0 - \sum_1^\infty b_n = \theta. \tag{14}$$

For all n, ϕ^{-1} is defined on x_n, and

$$\phi^{-1}(x_n) \leqq x_{n-1}. \tag{15}$$

From this it follows that ϕ^{-n} is defined on x_n and satisfies

$$\phi^{-n}(x_n) \leqq \phi^{-n+1}(x_{n-1}) \leqq \cdots \leqq \phi^{-1}(x_1) \leqq x_0, \tag{16}$$

and hence that

$$\phi^{-(n+1)}(x_n) \leqq \phi^{-(n-k+1)}(x_{n-k}) \leqq a_0 \qquad \text{for any } k < n. \tag{17}$$

But from (8) we have

(18) $$\phi^{-1}(b_n) = x_{n-1} - \phi^{-1}(x_n),$$

so we can define for $n \geqq 1$

(19) $$a_n \equiv \phi^{-(n+1)}(b_n) = \phi^{-n}(x_{n-1}) - \phi^{-(n+1)}(x_n) \leqq a_0.$$

It follows from (19) and (16) that for $k > 0$

(20) $$a_n a_{n+k} \leqq a_n \phi^{-(n+k)}(x_{n+k-1}) \leqq a_n \phi^{-(n+1)}(x_n) = \theta,$$

and hence that the elements a_n, $n = 1, 2, 3, \cdots$, are disjoint to each other. From this and from the fact that by definition $a_n \sim b_n$ (ϕ^{n+1}), we have

(21) $$\sum_1^\infty a_n \sim \sum_1^\infty b_n = b_0.$$

Since the element x is disjoint to both sides of (21), this yields

(22) $$\sum_1^\infty a_n + x_n \sim b_0 + x,$$

which with (12) gives

(23) $$\sum_1^\infty a_n + x \sim a_0 + x.$$

But by hypothesis $a + x$ is bounded, and so

(24) $$a_0 - \sum_1^\infty a_n = \theta.$$

We now have, by combining (21), and (24),

(25) $$a_0 \sim b_0,$$

which completes the proof of (4) and so of the required theorem.

4. **Consequences of Theorem** 3.2. We prove first the following statement.

4.1. THEOREM. *If a, b are bounded and if $a' \leqq a$ and $b' \leqq b$, then $a \sim b$ and $a' \sim b'$ implies that $a - a' \sim b - b'$; i.e., for bounded elements, Φ-equivalence is subtractive.*

Proof. Let $a \sim b$ (ϕ) and $a' \sim b'$ (ϕ'); then $a' \sim b' \sim \phi^{-1}(b') \leqq a$. Then by (3.2) $a - a' \sim a - \phi^{-1}(b') \sim b - b'$ (ϕ). Therefore $a - a' \sim b - b'$.

4.11. THEOREM. *If (a_n) and (b_n) are monotonic sequences of bounded elements of M and if $a_n \sim b_n$ for all n, then* $\lim a_n \sim \lim b_n$; *i.e., the respective limits of monotonic sequences of bounded equivalent elements are equivalent.*

Proof. Let us call these limits a and b, respectively. Suppose (a_n) and there-

fore also (b_n) are non-decreasing. Then $a=a_1+(a_2-a_1)+(a_3-a_2)+\cdots$, and $b=b_1+(b_2-b_1)+(b_3-b_2)+\cdots$. Then since a_n and b_n are bounded and $a_n\sim b_n$ for all n, by 4.1 we may sum corresponding sides and so get $a\sim b$. If the sequences are non-increasing, then from 4.1 $a_n-a_{n-1}\sim b_n-b_{n-1}$. Therefore $a_1-a\sim b_1-b$. Since further $a_1\sim b_1$, by again applying 4.1 we get $a\sim b$.

4.12. THEOREM. *The sum of any finite collection of bounded elements is bounded.*

Proof. We shall prove 4.12 for the case of two bounded, disjoint elements a_1 and a_2, since the general case is readily reducible to this. Suppose then that under these conditions a_1+a_2 is unbounded. Then there exists an element $a'\neq\theta$ contained in $a\equiv a_1+a_2$ such that $a-a'\neq\theta$, and $a\sim a'$. Then $a_1+a_2 \sim a_1'+a_2'$, where $a_1'=\phi(a_1)$ and $a_2'=\phi(a_2)$. From 3.2 we have

(1) $$a_1 - a_1a_1' \sim a_1' - a_1a_1',$$

(2) $$a_2 - a_2a_2' \sim a_2' - a_2a_2'.$$

But since $a_1'-a_1a_1'=a_1'a_2$ and $a_1-a_1a_1'=(a_1-a_1a')+a_1a_2'$ we have

(3) $$(a_1 - a_1a') + a_1a_2' \sim a_1'a_2,$$

and similarly

(4) $$(a_2 - a_2a') + a_2a_1' \sim a_2'a_1.$$

Now since the left-hand members of (3) and (4) are disjoint, we may substitute in (4) for $a_1'a_2$ and we get

(5) $$(a_2 - a_2a') + (a_1 - a_1a') + a_1a_2' \sim a_1a_2'.$$

But a_1a_2' is bounded and so

(6) $$(a_2 - a_2a') + (a_1 - a_1a') = a - a' = \theta,$$

contrary to hypothesis. Therefore a is bounded.

It may be noted that the class M of all bounded elements of M is a finitely additive ideal (and a restricted σ-ideal) in M, and that the set of all limits of (sequences of) bounded elements is a σ-ideal in M.

5. **Lattices M derived from rings of sets.** Let m be a σ-ring of subsets of a fixed set $\mathbf{s}$; n, a σ-ideal in m; and Π a class of univocal point transformations Π_α, defined between pairs of sets in m and leaving m and n invariant. Let M be the residue class ring m/n, and suppose the isomorphisms on M derived from the transformations of Π (see 1.22) satisfy (β). Suppose further that conditions (γ) hold:

5.1. i. *If $\mathbf{x}$ is any subset of $\mathbf{s}$ and if there exists a set $\mathbf{n}\in n$ such that $\mathbf{x}\subseteq\mathbf{n}$, then $\mathbf{x}\in n$.*

ii. *If $\mathbf{x}\subseteq\mathbf{s}$, then there exists a set $\mathbf{x}^*\in m$ such that* (a) $\mathbf{x}\subset\mathbf{x}^*$ *and* (b) *if $\mathbf{x}\subset\mathbf{a}\in m$, then $\mathbf{x}^*-\mathbf{x}^*\mathbf{a}\in n$.*

Let us denote the residue class of a set $\mathbf{x} \in m$ by $m(\mathbf{x})$ or by x'. Then

5.11. If $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ satisfy ii, then $x_1^* = x_2^*$.

5.12. If $\mathbf{x} \subseteq x'$, then $x^* \subseteq x'^*$.

5.13. For any $\mathbf{x} \subseteq \mathbf{s}$, there exists a set $\mathbf{x}_*$ of m, namely, $\mathbf{x}^* - (\mathbf{x}^* - \mathbf{x})^*$, such that i. $\mathbf{x}_* \subseteq \mathbf{x}$; ii. if $\mathbf{x}_*' \subset \mathbf{x}$, then $\mathbf{x}_*' \subseteq \mathbf{x}_*$, and iii. if $\mathbf{x}_*$ and $\mathbf{y}_*$ satisfy i and ii, then $x_* = y_*$.

5.14. *A set $\mathbf{x} \subset \mathbf{s}$ is in m if and only if $x^* = x_*$, in which case $x^* = x_* = x$.*

5.15. If $\mathbf{x}_n \subset \mathbf{s}$, then for any $\mathbf{x}_n^*$, $\mathbf{x}_{n*}$, $m(\sum \mathbf{x}_n^*) = \sum m(\mathbf{x}_n^*)$; $m[(\sum \mathbf{x}_n)_*] \geqq m(\sum \mathbf{x}_{n*})$; $m[(\prod \mathbf{x}_n)^*] \leqq m(\prod \mathbf{x}_n^*)$; $m[(\prod \mathbf{x}_n)_*] = m(\prod \mathbf{x}_{n*})$.

Since the proofs of these relations are all alike, we shall prove only the first. From the fact that $(\sum \mathbf{x}_n)^* \geqq \sum \mathbf{x}_n > \mathbf{x}$ for all n, $m(\mathbf{x}_n^*) \leqq m[(\sum \mathbf{x}_n)^*]$, and so $m[(\sum \mathbf{x}_n)^*] \geqq m[\sum \mathbf{x}_n^*]$. Similarly, $\mathbf{x}_n^* \geqq \mathbf{x}_n$ for all n, and so $\sum \mathbf{x}_n^* \geqq \sum \mathbf{x}_n$. Therefore $m(\sum x_n^*) > m[(\sum \mathbf{x}_n)^*]$, and the relation is proved.

5.16. Lemma. *A necessary and sufficient condition that $\mathbf{x} \in m$ is that for any $y \subseteq \mathbf{s}$, $(\mathbf{x}y)^*(y - \mathbf{x}y)^* \in n$.*

Proof. Suppose $\mathbf{x} \notin m$. Then for any $\mathbf{x}^*$, $\mathbf{x}^* - \mathbf{x}$ is not in m and so not in n. Then from i of (γ), $(\mathbf{x}^* - \mathbf{x})^*$ is not in n. Now let $y = \mathbf{x}^*$. Then $(\mathbf{x}y)^*(\mathbf{x} - \mathbf{x}y)^* = (\mathbf{x}^* - \mathbf{x})^* \notin n$. If $\mathbf{x} \in m$, then for any $y \subseteq \mathbf{s}$, $(\mathbf{x}y)^* \leqq \mathbf{x}$, and so $m[(\mathbf{x}y)^*] \leqq m(\mathbf{x})$. Also $y - \mathbf{x}y < \mathbf{s} - \mathbf{x} < (\mathbf{s} - \mathbf{x})^* - (\mathbf{s} - \mathbf{x})^*(\mathbf{x}y)^*$, which with ii of (γ) completes the proof.

5.17. If π is any transformation of Π defined on a set $\mathbf{a} \in m$, then for any $\mathbf{x} \subset \mathbf{a}$, we define $\pi(\mathbf{x})$ as the image of $\mathbf{x}$ under π. Then since π leaves m and n invariant, $\pi(\mathbf{a}) \in m$, and $\pi(\mathbf{a}) \supseteq \pi(\mathbf{x}^*) \supseteq \pi(\mathbf{x})$, so $m[\pi(\mathbf{x}^*)] > m[\pi^*(\mathbf{x})]$. But since $\pi^*(\mathbf{x}) \supseteq \pi(\mathbf{x})$, it follows that $\pi^{-1}[\pi^*(\mathbf{x})] \supseteq \mathbf{x}$. Therefore $\pi^{-1}[\pi^*(\mathbf{x})] \supseteq \mathbf{x}^*$, whence $m[\pi(\mathbf{x}^*)] = m[\pi^*(\mathbf{x})]$. By a similar argument with reference to $\mathbf{x}$, we also have $m[\pi(\mathbf{x}_*)] = m[\pi_*(\mathbf{x})]$. *That is, $m(\mathbf{x}^*) \sim m[\pi(\mathbf{x})^*]$, and $m(\mathbf{x})_* \sim m[\pi(\mathbf{x})_*]$.*

Part II

6. **Measuring systems in abstract sets.** A set $\mathbf{s}$ is said to have a *measuring system* (m, n, Φ) if m is a σ-ring of subsets of $\mathbf{s}$; n, a σ-ideal in m (notice that under these conditions $M \equiv m/n$ satisfies (α) of §1); and Φ, a class of σ-isomorphisms defined on M and satisfying (β) (§2). If (γ) of §5 also holds, the system (m, n, Φ) is *complete*. Any set has a trivial measuring system, namely, (the totality of its subsets, the void set, the identity transformations on every subset). In euclidean space, the Lebesgue measurable sets, the null sets, and the transformations derived from direct sums of rigid motions form such a system.

Now let us consider a set $\mathbf{s}$ with a given system (m, n, Φ). An isomorphism of Φ is said to be *measure-preserving*, and the elements $a, b, c, \cdots$, of M to be *measurable* and of *measure* $M(a)$, $M(b)$, $M(c)$, $\cdots$, (2.2, i), respectively. Henceforth these measures will be written $\mathfrak{a}$ or $\mathfrak{m}(a)$, $\mathfrak{b}$ or $\mathfrak{m}(b)$, $\mathfrak{c}$ or $\mathfrak{m}(c)$, $\cdots$. In particular, the element n is said to have null measure $\mathfrak{m}(n) = 0$. The ele-

ments a and b are of equal measure if and only if they correspond under some measure-preserving transformation, i.e., if and only if $a \sim b$. If the system (m, n, Φ) is complete, a set $\mathbf{a} \in m$ is said to be *measurable* and of *measure* $\mathfrak{m}(m(\mathbf{a}))$, where $m(\mathbf{a})$ is the residue class of $\mathbf{a}$. For any $\mathbf{x} \subseteq \mathbf{s}$, the elements x^* and x_* (§5) are called a *hull* and a *kernel* of $\mathbf{x}$, respectively. Then $\mathfrak{m}(x^*)$ and $\mathfrak{m}(x_*)$, which will now be written $\mathfrak{x}^*$ and $\mathfrak{x}_*$, will be called the *outer measure* and the *inner measure* of $\mathbf{x}$.

7. **Partial ordering of the measure-values. Addition.** Let $\mathfrak{M}^* = (\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \cdots)$ be the totality of measures, and let $\mathfrak{M}$ be the subset of bounded measures (2.32). Henceforth $\mathfrak{a} < \infty$ shall mean $\mathfrak{a}$ is bounded.

7.1. Definition. *If $\mathfrak{a} < \infty$, we define $\mathfrak{a} < \mathfrak{b}$ to mean that there exist corresponding elements a and b in M such that $a < b$. It follows that $\mathfrak{a} < \mathfrak{b}$ is a (proper) partial ordering of $\mathfrak{M}$. The null element* 0 *of $\mathfrak{M}^*$ is the zero of the ordering.*

7.11. Any monotonic increasing (decreasing) sequence $(\mathfrak{a}_n)$ has a supremum (an infimum) written $\lim \mathfrak{a}_n$. For suppose $\mathfrak{a}_n$ increasing, and let a_n be a sequence of corresponding elements from M. For each n, there exists an $a_n' < a_{n+1}$ such that $a_n' \sim a_n$. Then $\mathfrak{m}[\sum(a_{n+1} - a_n')]$ can be verified to be the desired supremum.

7.2. Definition. *If $(\mathfrak{a}_n) \subset \mathfrak{M}^*$, and if there exists a sequence (a_n) of disjoint corresponding elements in M, then $(\mathfrak{a}_n)$ is said to have a sum $\sum \mathfrak{a}_n$, namely, $\mathfrak{m}(\sum a_n)$. A binary sum will be written $\mathfrak{a} + \mathfrak{b}$.*

It follows from the considerations of Part I that:

7.21. Whenever it exists, the sum is unique, and the operation of addition is commutative and associative, and $\sum \mathfrak{a}_n = \lim \sum_1^n \mathfrak{a}_i$.

7.22. If $\mathfrak{a}_i = \mathfrak{a}$ for all i, then we write $\sum_1^n \mathfrak{a}_i$ as $n\mathfrak{a}$. If either side of the following exists, so does the other; and $n\mathfrak{a} + m\mathfrak{a} = (n+m)\mathfrak{a}$; $n(\mathfrak{a}+\mathfrak{b}) = n\mathfrak{a} + n\mathfrak{b}$; $n(m\mathfrak{a}) = m(n\mathfrak{a}) = nm\mathfrak{a}$.

7.23. For any $\mathfrak{a} \in \mathfrak{M}^*$, $\mathfrak{a} + 0$ exists and is equal to $\mathfrak{a}$; if $\mathfrak{a} < \infty$ then $\mathfrak{a} + \mathfrak{b} = \mathfrak{a}$ if and only if $\mathfrak{b} = 0$.

7.24. Binary addition has a unique inverse for bounded elements; i.e., $\mathfrak{a} + \mathfrak{x} = \mathfrak{b} < \infty$ has a unique solution for fixed $\mathfrak{a}$ and $\mathfrak{b}$, denoted by $\mathfrak{b} - \mathfrak{a}$.

7.25. A necessary and sufficient condition that $\mathfrak{a} < \mathfrak{b}$ is that for some $\mathfrak{c}$, $\mathfrak{a} + \mathfrak{c} = \mathfrak{b}$ holds. If $\mathfrak{a} < \mathfrak{b} < \infty$ then there exists a maximum integer n such that $n\mathfrak{a} \leqq \mathfrak{b}$.

7.26. If $\sum \mathfrak{m}(a_n)$ exists, then $\sum \mathfrak{m}(a_n) \geqq \mathfrak{m}(\sum a_n)$, and equality holds if and only if the elements a_n are disjoint, in which case $\sum \mathfrak{m}(a_n)$ does exist.

8. **Algebraic properties of $\mathfrak{M}^*$.**

8.1. We now see that any measuring system (m, n, Φ) gives rise to a set $\mathfrak{M}^*$ of measure-values with a commutative and associative operation of addition defined for certain denumerable sets of elements of $\mathfrak{M}^*$, in terms of which a partial ordering of the bounded elements can be derived. Then

i. *There exists an element* 0 *such that* $\mathfrak{a}\neq 0$ *implies that* $\mathfrak{a}>0$ *and* $\mathfrak{a}+0=\mathfrak{a}$.

ii. *If* $a<\infty$, *and* $a<b$, *then* $\mathfrak{m}(a)<\mathfrak{m}(b)$.

iii. $\sum \mathfrak{m}(a_n)\geqq \mathfrak{m}(\sum a_n)$ *and equality holds if and only if the elements* a_n *are disjoint.*

If, further, the measure defined is complete, then to any $\mathbf{x}\subseteq \mathbf{s}$ there corresponds an outer and an inner measure $\mathfrak{x}^*$ and $\mathfrak{x}_*$, respectively, and

iv. *If* $x^*<\infty$, *and* $\mathbf{x}<\mathbf{y}$, *then* $\mathfrak{x}^*\leqq \mathfrak{y}^*$.

v. *If* (x_n) *is a sequence of sets such that* $\sum x_n^*$ *exists, then* $\mathfrak{m}(\sum x_n)^*\leqq \sum \mathfrak{x}_n^*$ *if* $(\sum \mathbf{x}_n)^*<\infty$.

vi. *The set* $\mathbf{x}$ *is measurable if and only if the equivalent conditions* a *and* b (*the Carathéodory condition*) *hold*:

a. $x^*=x_*$, *in which case* $x^*=x_*=x$, *and*

b. *for any* $\mathbf{y}\subseteq \mathbf{s}$, $\mathfrak{m}(xy)^*+\mathfrak{m}(y-xy)^*=y^*$ $[=\mathfrak{m}((xy)^*\cdot(y-xy)^*)]$.

9. **Ordered sets of measure-values.** In the remainder of Part II it is assumed that the system (m, n, Φ) satisfies condition (δ).

(δ) *For any* a, b *of* M *either there exists a* $b'\leqq b$ *such that* $a\sim b'$ *or there exists an* $a'<a$ *such that* $a'\sim b$.

9.1. *A necessary and sufficient condition that* (δ) *hold is that for any two elements* $\mathfrak{a}$ *and* $\mathfrak{b}$ *of* $\mathfrak{M}^*$ *either there exists an* $\mathfrak{a}'$ *such that* $\mathfrak{a}+\mathfrak{a}'=\mathfrak{b}$ *or a* $\mathfrak{b}'$ *such that* $\mathfrak{b}+\mathfrak{b}'=\mathfrak{a}$.

We shall show that (δ) is a sufficient as well as a trivially necessary condition that M be order and addition isomorphic to a set of positive numbers.

9.2. For any bounded measures $\mathfrak{a}$ and $\mathfrak{b}$, $\mathfrak{a}\nless\mathfrak{b}$ implies that $\mathfrak{a}\geqq\mathfrak{b}$; i.e., $\mathfrak{M}$ is linearly ordered, and either $\mathfrak{b}-\mathfrak{a}$ or $\mathfrak{a}-\mathfrak{b}$ exists.

9.21. From 9.2, one and only one of the following occurs: a. $\mathfrak{M}$ contains the single element 0; b. there exists an element $\mathfrak{a}$ of $\mathfrak{M}$ such that $\mathfrak{a}\neq 0$, and if $\mathfrak{a}'<\mathfrak{a}$, then $\mathfrak{a}'=0$; c. for any $\mathfrak{a}>0$ there exists an $\mathfrak{a}'$ such that $0<\mathfrak{a}'<\mathfrak{a}$. In the last case, we may assume that $2\mathfrak{a}'\leqq\mathfrak{a}$, for we can take the lesser of $\mathfrak{a}'$ and $\mathfrak{a}-\mathfrak{a}'$.

9.22. If $\mathfrak{a}_n$, $\mathfrak{a}$ are bounded, and $\mathfrak{a}_n\leqq\mathfrak{a}_{n+1}(\mathfrak{a}_n\geqq\mathfrak{a}_{n+1})$, then $\mathfrak{a}<\lim \mathfrak{a}_n$ $(\mathfrak{a}>\lim \mathfrak{a}_n)$ implies that there exists a k such that $\mathfrak{a}\leqq\mathfrak{a}_k$ $(\mathfrak{a}\geqq\mathfrak{a}_k)$.

9.3. i. For any $\mathfrak{a}$, $\mathfrak{b}$ in $\mathfrak{M}$, there is a unique integer n such that $\mathfrak{b}=n\mathfrak{a}+\mathfrak{c}$, $\mathfrak{c}<\mathfrak{a}$. ii. If $\mathfrak{a}$ and $\mathfrak{b}$ are bounded, then for every integer n such that $n\mathfrak{a}$ and $n\mathfrak{b}$ exist, $n\mathfrak{a}$ and $n\mathfrak{b}$ stand in the same relative order as do $\mathfrak{a}$ and $\mathfrak{b}$.

9.4. There exists an element s of M^* such that $\mathfrak{a}<\infty$ implies that $\mathfrak{a}\leqq\mathfrak{s}$. If $\mathfrak{s}\nless\infty$, the sum of any two bounded elements always exists; if $\mathfrak{s}<\infty$, $\mathfrak{a}+\mathfrak{b}$ exists if and only if the equivalent conditions $\mathfrak{a}<\mathfrak{s}-\mathfrak{b}$ and $\mathfrak{b}<\mathfrak{s}-\mathfrak{a}$ hold. In this latter case, the element s is unique; and for any a of M, $a\leqq s$.

9.5. If $\mathfrak{M}$ satisfies b of 9.21, that is, if there is a least element $\mathfrak{a}_1>0$, then it follows from 9.3 and 9.4 that the elements of $\mathfrak{M}$ are of the form $n\mathfrak{a}_1$, and that according as $\mathfrak{s}$ is unbounded or bounded and equal to $k\mathfrak{a}_1$, n will assume all positive integral values or all those not greater that k.

9.6. Lemma. *If* $\mathfrak{M}$ *satisfies* c *of* 9.21, *i.e.*, $\mathfrak{M}$ *has no minimal non-null ele-*

ments, and, if each element of $\mathfrak{M}$ is contained in one or the other, but not both of two non-void subclasses $\mathfrak{M}'$ and $\mathfrak{M}''$ and if every element of $\mathfrak{M}''$ is greater than every element of $\mathfrak{M}'$, then either there is a least element of $\mathfrak{M}''$ and no greatest in $\mathfrak{M}'$, or there is a greatest element in $\mathfrak{M}'$ and no least in $\mathfrak{M}''$.

Proof. Since neither $\mathfrak{M}'$ nor $\mathfrak{M}''$ is void there exist elements $\mathfrak{a}_1'$ in $\mathfrak{M}'$ and $\mathfrak{a}_1''$ in $\mathfrak{M}''$ with $\mathfrak{a}_1' < \mathfrak{a}_1''$. From c, there is a third element between the two, so both cannot be the extreme elements of their respective classes. If either one is, the lemma is proved; if not, since there is some $\mathfrak{c}_1 > 0$ such that $2\mathfrak{c}_1 \leqq \mathfrak{a}_1'' - \mathfrak{a}_1'$, there is a maximum integer n_1 such that $\mathfrak{a}_1' + n_1\mathfrak{c}_1 < \mathfrak{M}'$, and a maximum integer $n_1' \geqq n_1$ such that $\mathfrak{a}_1' + n_1'\mathfrak{c}_1 \leqq \mathfrak{a}_1''$. Let $\mathfrak{a}_2' = \mathfrak{a}_1' + n_1\mathfrak{c}_1$ and let $\mathfrak{a}_2''$ be either $\mathfrak{a}_1' + (n_1' + 1)\mathfrak{c}_1$ or $\mathfrak{a}_1''$, according as $n_1' > n_1$ or $n_1' = n_1$. Then $\mathfrak{a}_2' \in \mathfrak{M}'$, $\mathfrak{a}_2'' \in \mathfrak{M}''$, $\mathfrak{a}_1' \leqq \mathfrak{a}_2' < \mathfrak{a}_2'' \leqq \mathfrak{a}_1''$, and $2(\mathfrak{a}_2'' - \mathfrak{a}_2') \leqq (\mathfrak{a}_1'' - \mathfrak{a}_1')$. Let $\mathfrak{c}_2$ be the minimum of $\mathfrak{a}_1'' - n_1'\mathfrak{c}_1$ and $\mathfrak{c}_1 - (\mathfrak{a}_1'' - n_1'\mathfrak{c}_1)$. Then $2\mathfrak{c}_2 \leqq \mathfrak{c}_1$ and hence $4\mathfrak{c}_2 \leqq \mathfrak{a}_1'' - \mathfrak{a}_1'$. Either if $\mathfrak{a}_2'$ is the maximum of its class, or if $\mathfrak{a}_2''$ is the minimum of its class, the assertion holds. If not, we define $\mathfrak{a}_3'$ and $\mathfrak{a}_3''$ just as in the preceding case. By proceeding in this way, we arrive at a pair of sequences $(\mathfrak{a}_n')$ and $(\mathfrak{a}_n'')$ of elements of $\mathfrak{M}'$ and $\mathfrak{M}''$, respectively, such that $\mathfrak{a}_n' \leqq \mathfrak{a}_{n+1}' \leqq \mathfrak{a}_{n+1}'' \leqq \mathfrak{a}_n''$ and $2^{n-1}(\mathfrak{a}_n'' - \mathfrak{a}_n') \leqq \mathfrak{a}_1'' - \mathfrak{a}_1'$. The sequences are monotonic and contained in M, so their respective limits $\mathfrak{a}'$ and $\mathfrak{a}''$ exist and are bounded. Also, since for any $\mathfrak{b}' \in \mathfrak{M}'$ and for any integer m, $\mathfrak{b}' < \mathfrak{a}_m''$, it follows from 7.11 that $\mathfrak{b}' \leqq \mathfrak{a}'' = \lim \mathfrak{a}_n''$. In particular this relation holds for all elements of $(\mathfrak{a}_n')$, and so $\mathfrak{a}' = \lim \mathfrak{a}_n' \leqq \mathfrak{a}''$. Further since $\mathfrak{a}_n' \leqq \mathfrak{a}' \leqq \mathfrak{a}'' \leqq \mathfrak{a}_n''$ for all n, it follows that $\mathfrak{a}'' - \mathfrak{a}' \leqq \mathfrak{a}_n'' - \mathfrak{a}_n' \leqq 2^{-n-1}(\mathfrak{a}_1'' - \mathfrak{a}_1')$ and hence that $\mathfrak{a}'' - \mathfrak{a}' = \theta$, i.e., $\mathfrak{a}' = \mathfrak{a}'' = \mathfrak{a}$. From the fact that $\mathfrak{a}$ is the limit of a sequence from $\mathfrak{M}'$ and of one from $\mathfrak{M}''$ it follows that $\mathfrak{a}$ is the greatest or the least element of its class according as it is in $\mathfrak{M}'$ or in $\mathfrak{M}''$, and that the class to which $\mathfrak{a}$ does not belong has no extreme element.

10. Isomorphisms of $\mathfrak{M}$ onto sets of positive numbers.

10.1. THEOREM. *Under condition* (δ) *the set $\mathfrak{M}$ of bounded measure-values is isomorphic to a subset Γ of the set of non-negative numbers; i.e., there exists a one-to-one mapping $f(\mathfrak{a})$ of $\mathfrak{M}$ onto Γ such that whenever $\sum \mathfrak{a}_n$ exists and is bounded then $\sum f(\mathfrak{a}_n)$ converges to $f(\sum \mathfrak{a}_n)$, and whenever $\sum \mathfrak{a}_n$ exists and is unbounded, then $\sum f(\mathfrak{a}_n)$ diverges to infinity. The mapping f is completely determined up to a multiplicative constant; that is, if f' is any other mapping of M onto Γ with the above property, then there exists a constant ξ such that for all $a < \infty$, $f(\mathfrak{a}) = \xi f'(\mathfrak{a})$. According as conditions* a, b, *or* c, *of* 9.21 *hold, the set Γ will consist of* a. *the number zero,* b. *the set $nf(\mathfrak{a}_1)$, where $n = 1, 2, 3, \cdots$ or $1, 2, \cdots, k$, or* c. *the interval $0 \leqq \xi \leqq f(\mathfrak{s})$, or $0 \leqq \xi \leqq \infty$, according as the element $\mathfrak{s}$ of 9.4 is bounded or unbounded.*

The theorem follows immediately for cases a and b. In case c, we shall use the following lemma:

LEMMA. *To every ordered pair* $(\mathfrak{a}, \mathfrak{b})$ *of bounded elements different from* 0, *there corresponds a unique positive finite number* $\mu(\mathfrak{a}, \mathfrak{b})$ *such that*

i. *if* $\sum \mathfrak{a}_n$ *exists, then* $\mu(\sum \mathfrak{a}_n, \mathfrak{b}) = \sum \mu(\mathfrak{a}_n, \mathfrak{b})$,

ii. $\mu(\mathfrak{a}, \mathfrak{a}) = 1$.

Proof. For any $\mathfrak{a} \neq 0$, the set $\mathfrak{M}' \subseteq \mathfrak{M}$ of elements $\mathfrak{a}'$ such that $2\mathfrak{a}' \leqq \mathfrak{a}$, and the set $\mathfrak{M}'' \subseteq \mathfrak{M}$ of elements $\mathfrak{a}''$ such that $2\mathfrak{a}'' > \mathfrak{a}$, satisfy the hypotheses of 9.6. From this it follows that there exists an element $\mathfrak{a}^1$ such that $2\mathfrak{a}^1 = \mathfrak{a}$, and such that for every integer n an $\mathfrak{a}^n$ exists such that $2^n\mathfrak{a}^n = \mathfrak{a}$. Then

$$\mathfrak{a}^n = 2^k\mathfrak{a}^{n+k}, \tag{1}$$

$$\sum_{i=1}^{p} \mathfrak{a}^{n_i} = \left(\sum_{1}^{p} 2^{n_i - m} \right) \mathfrak{a}^m \qquad \text{if } n_i > m \text{ for } i = 1, 2, \cdots, p. \tag{2}$$

Further if $\mathfrak{a}' > 0$, there exists an n such that $\mathfrak{a}' > \mathfrak{a}^n$. For by 9.4 there exists a k such that $\mathfrak{a} = k\mathfrak{a}' + \mathfrak{a}''$, where $\mathfrak{a}'' < \mathfrak{a}'$, and we may take $n = k+1$. More generally,

$$\text{if } \mathfrak{a}' > \mathfrak{a}'', \text{ there exists an } n \text{ for which } \mathfrak{a}' < \mathfrak{a}'' + \mathfrak{a}^n. \tag{3}$$

For every pair of elements $\mathfrak{a}$ and $\mathfrak{b}$ there exists an integer $\mu_n(\mathfrak{a}, \mathfrak{b})$ such that $\mathfrak{a} = \mu_n \mathfrak{b}^n + \mathfrak{c}_n$, where $\mathfrak{c}_n < \mathfrak{b}^n$. From the previous, it follows that the μ_n satisfy

$$2\mu_n \leqq \mu_{n-1} \leqq 2(\mu_n - 1); \tag{4}$$

(5) if $\mathfrak{c} > \mathfrak{a}$, there exists a k such that for all $p > 0$, $\mu_{k+p}(\mathfrak{c}, \mathfrak{b}) > \mu_{k+p}(\mathfrak{a}, \mathfrak{b})$;

$$\mu_k \left(\sum_{i=1}^{p} \mathfrak{a}^{n_i}, \mathfrak{a} \right) = \sum_{1}^{p} \mu_k(\mathfrak{a}^{n_i}, \mathfrak{a}) \qquad \text{for } n_i > k \text{ for } i = 1, 2, \cdots, p; \tag{6}$$

$$\mu_r(\mathfrak{b} + k\mathfrak{a}^r, \mathfrak{a}) = \mu_r(\mathfrak{b}, \mathfrak{a}) + k/2^r, \qquad \text{for any } \mathfrak{b}. \tag{7}$$

From (4) the sequence of numbers $\mu_n(\mathfrak{b}, \mathfrak{a})$ is monotonically increasing and bounded and therefore approaches a unique limit $\mu(\mathfrak{b}, \mathfrak{a})$. We have at once that

$$\mu(0, \mathfrak{a}) = 0; \qquad \mu(\mathfrak{a}_n, \mathfrak{a}) = 1/2^n; \tag{8}$$

$$\text{if } \mathfrak{c} < \mathfrak{b}, \text{ then } \mu(\mathfrak{c}, \mathfrak{a}) < \mu(\mathfrak{b}, \mathfrak{a}); \tag{9}$$

$$\mu(\mathfrak{c} + k\mathfrak{a}^r) = \mu(\mathfrak{c}) + k/2^r. \tag{10}$$

Also for any finite p, $\mu(\sum_1^p \mathfrak{a}_i, \mathfrak{a}) = \sum_1^p \mu(\mathfrak{a}_i, \mathfrak{a})$. For $\sum_1^p \mathfrak{a}_i = \sum_1^p \mu_k(\mathfrak{a}', \mathfrak{a})\mathfrak{a}^k + \sum_{i=1}^p \mathfrak{a}'_{k,i}$, where $\mathfrak{a}'_{k,i} < \mathfrak{a}^k$. It can be seen that $\mu_k(\sum_{i=1}^p \mathfrak{a}_i, \mathfrak{a}) < p$; then applying (10), dividing by 2^k and passing to the limit give the desired result. Now suppose that $\mathfrak{a}^* = \sum_1^\infty \mathfrak{a}_i$ exists. If $\mathfrak{a}^* < \infty$, then since for every n $\mu_n(\mathfrak{a}^*, \mathfrak{a})\mathfrak{a}^n \leqq \mathfrak{a}^*$ it follows that there is a p_n for which $\mu_n(\mathfrak{a}^*, \mathfrak{a})\mathfrak{a}^n \leqq \sum_1^p \mathfrak{a}_i$. Then $\mu_n(\mathfrak{a}^*, \mathfrak{a})/2^n = \mu[\mu_n(\mathfrak{a}^*, \mathfrak{a})\mathfrak{a}^n] \leqq \mu(\sum_{i=1}^p \mathfrak{a}_i, \mathfrak{a}) < \mu(\mathfrak{a}^n, \mathfrak{a})$, or $\mu(\mathfrak{a}^*, \mathfrak{a}) - \mu(\sum_1^p \mathfrak{a}_i, \mathfrak{a}) < 1/2^n$. If $\sum_1^\infty \mathfrak{a}_i \not< \infty$, then for any n, $n\mathfrak{a} < \sum_1^\infty \mathfrak{a}_i$, whence $\mu(\sum \mathfrak{a}_i, \mathfrak{a}) > \mu(n\mathfrak{a}, \mathfrak{a}) \geqq n$.

Complete additivity of $\mu(\mathfrak{b}, \mathfrak{a})$ follows from these last two relations and the finite additivity; i.e.,

$$\mu(\sum \mathfrak{a}_i, \mathfrak{a}) = \sum \mu(\mathfrak{a}_i, \mathfrak{a}). \tag{11}$$

Relations i and ii of the lemma are now proved, and it remains to show for each ordered pair $\mathfrak{b}$, $\mathfrak{a}$ that $\mu(\mathfrak{b}, \mathfrak{a})$ is unique. But any $\mu'(\mathfrak{b}, \mathfrak{a})$ satisfying i and ii can be seen to be such that

$$k\mu'(\mathfrak{a}^n, \mathfrak{a}) = k/2^n = \mu(\mathfrak{a}^n, \mathfrak{a}), \qquad \text{for all } k \text{ and } n; \tag{12}$$

and so

$$\mu'(\mathfrak{b}, \mathfrak{a}) = \mu'(\lim \mu_n(\mathfrak{b}, \mathfrak{a})\mathfrak{a}^n) = \lim \mu'(\mu_n(\mathfrak{b}, \mathfrak{a})\mathfrak{a}^n) = \lim \mu_n/2^n = m(\mathfrak{b}, \mathfrak{a}). \tag{13}$$

We can also verify by a limiting process that if none of $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$ is 0, then

$$\mu(\mathfrak{a}, \mathfrak{b}) = \mu(\mathfrak{a}, \mathfrak{c})\mu(\mathfrak{c}, \mathfrak{b}). \tag{14}$$

Then the isomorphism required in 10.1 can be given by $f(\mathfrak{a}) = \mu(\mathfrak{a}, \mathfrak{a}_1)$ for any $\mathfrak{a}_1 > 0$. From (14), $f(\mathfrak{a})/f(\mathfrak{b}) = \mu(\mathfrak{a}, \mathfrak{a}_1)/\mu(\mathfrak{b}, \mathfrak{a}_1) = \mu(\mathfrak{a}, \mathfrak{b})$, which is independent of f and of $\mathfrak{a}_1$. But if $g(\mathfrak{a})$ is any mapping whatever satisfying the conditions of 10.1, then the ratio $g(\mathfrak{a})/g(\mathfrak{b})$ satisfies the conditions of the lemma, and so is equal to $\mu(\mathfrak{a}, \mathfrak{b})$. Therefore $g(\mathfrak{a}) = g(\mathfrak{a}_1)\mu(\mathfrak{a}, \mathfrak{a}_1)$, and $g(\mathfrak{a})/f(\mathfrak{a}) = g(\mathfrak{a}_1)$, a constant.

Let ξ be any positive finite number less than $\mu(\mathfrak{s}, \mathfrak{a})$, where $\mathfrak{a} \neq 0$, and $\mathfrak{s}$ satisfies 9.4. Then ξ can be expressed uniquely in the form $\zeta + \xi/2^k$ where ζ is an integer, $\xi_k = 0$ or 1 and an infinite set of the ξ_k are zero. It follows from 9.3 and relation (11) that all the sums $\xi_k \mathfrak{a}^k$ exist and that $\mu(\sum \xi_k)\mathfrak{a} = \xi$. Therefore the range of any isomorphism, which must be of the form $f(\mathfrak{b}) = \mu(\mathfrak{b}, \mathfrak{a})$, is a closed finite or infinite interval with left end point zero. This completes the proof of 10.1.

10.2. Let s' and s'' be any two unbounded elements of M, if such exist, which are the respective limits of monotonic increasing sequences (s_n') and (s_n'') of bounded elements. Then there exist subsequences (s_{n_i}') and (s_{n_i}'') such that $s_{n_i} \leqq s_{n_i}'' < s'_{n_{i+1}}$. It follows that $\mathfrak{s}' = \mathfrak{s}''$, that is, all unbounded limits of bounded elements of M are of equal measure, which we denote by $+\infty$. If the system (m, n, Φ) is such that

(ϵ) *every element of M is the limit of a sequence of bounded elements*,

then the range of measure-values in $\mathfrak{M}^*$ is $0 \leqq \xi \leqq +\infty$.

11. **Isomorphisms of $\mathfrak{M}$ onto the infinite half-interval. Conclusion.** We have shown that for any set $\mathfrak{s}$, any system (m, n, Φ) satisfying (α) and (β) leads to a set $\mathfrak{M}^*$ of measures, the bounded subset $\mathfrak{M}$ of which satisfies conditions which may be regarded as a generalization of the Lebesgue conditions on a numerical measure. If in addition (δ) holds, the set $\mathfrak{M}^*$ of limits of bounded measures is isomorphic to either a set of successive non-negative integers or to a closed interval with left end points zero. If (ϵ) holds, i.e., if

$\mathfrak{M}_* = \mathfrak{M}^*$, the measure determined by (m, n, Φ) can be represented by a function $f(a)$ which is unique up to a multiplicative constant, assumes the values $0 \leqq \xi \leqq \infty$, and is such that $f(a) = 0$ if and only if $a = \theta$; $f(a) = f(b)$ if and only if a and b correspond under some measure-preserving transformation, and $f(a) < \infty$ if and only if $a < \infty$. It may be pointed out that no reference is made to any topological or metric properties of the set $\mathbf{s}$.

It is not necessarily true that every numerical measure can be derived by the procedure above, for this would require the existence of a measure-preserving transformation between any two sets of equal measure. However, it is shown in Part III that this is true for many important measures. The general case can be treated by introducing into M an equivalence relation, equality of measure, which is countably disjointly additive, hereditary in the sense of 2.2, and for bounded elements is also subtractive and preserved by taking limits of monotonic sequences. It can easily be verified that the entire theory of Part II would hold in this case. Further if M has a countable basis, a family of isomorphisms can be defined which generate the given congruence, which reduces this case to the previous one.

Part III

12. **Lattices satisfying (ζ). Bases.** Let M be a lattice satisfying (α) of §1 and $|a|$ be a functional defined on M and satisfying (ζ):

ia. $0 \leqq |a| \leqq \infty$, *for any a of M, and $|a| = 0$ if and only if $a = \theta$;*

ib. *if $|a| = \infty$, then there exists a sequence of elements a_n such that $|a| < \infty$, and $\sum a_n = a$;*

ii. $|\sum a_n| = \sum |a_n|$, *for any sequence of disjoint a_n;*

iii. *for any $a \neq \theta$ there is an $a' < a$ such that $0 \leqq |a| \leqq \infty$;*

iv. *for any a of M, there is a countable set of elements a_n of M such that for any $a' < a$, and for any $\epsilon > 0$, some subsequence (a_m) covers a' and $|\sum a_m - a'| < \epsilon$.*

The object of the next sections is to prove under conditions (ζ) that if $|a| = |a'|$ then there exists a σ-isomorphism ϕ of $I(a)$ onto $I(a')$ (see 1.1) such that for any $b < a$, $|\phi(b)| = |b|$, which in turn leads to the result that the $|a|$ can be determined up to a constant multiple by a set of isomorphisms satisfyig (β).

Henceforth a family of elements a_n satisfying iv of (ζ) will be called a *basis* of a, and any element of a basis a *base element.*

12.1. From the fact that if each element of a basis can be expressed as the sum of elements of a family (a_n') then (a_n') is also a basis, it follows that a has a basis of the following form:

Let (b_n) be any basis of a. Then let a_n be the nth element of the form $b_k - \sum_1^{k-1} b_i$ which is different from θ; let $a_{n,0}$ be the first element of the form $a_n b_{k+i}$, where a_n is derived from b_k such that $\theta < a_n b_{k+i} < a_n$; and let $a_{n,1} = a_n - a_{n,0}$. Then if a_ν is defined for every sequence $\nu = (n, \epsilon_1, \cdots, \epsilon_k)$, $\epsilon_i = \pm 1$, as some product $\epsilon_1 b_1 \cdots \epsilon_k b_k$ where $-b = a - b$, then $a_{\nu,0}$ is given

by $\epsilon_1 b_1 \cdots \epsilon_k b_k b_m$, where m is the least value for which $0<\epsilon_1 b_1 \cdots \epsilon_k b_k b_m <\epsilon_1 b_1 \cdots \epsilon_k b_k$, and $a_{\nu,1}=a_\nu - a_{\nu,0}$. It follows at once that every b_n is the sum of a finite set of the a_ν. Therefore (a_ν) is a basis. Let $n(\nu)$ be the number of terms in the sequence ν, and Γ_k the totality of sequences ν such that $n(\nu)=k$. We define $\gamma<\nu$, where $\gamma=(n_1, \epsilon_1, \cdots, \epsilon_k)$ and $\nu=(n', \epsilon_1', \cdots, \epsilon_k')$ if $k>k'$, and $(n, \epsilon_1, \cdots, \epsilon_k)\equiv(n', \epsilon_1', \cdots, \epsilon_k')$.

12.2. Then

i. $\sum a_\gamma = a$, $\gamma\in\Gamma_n$, for any fixed n;

ii. $a_\nu a_\gamma=\theta$ if $n(\nu)=n(\gamma)$;

iii. $a_\nu<a_\gamma$ if and only if $\nu<\gamma$;

iv. $a_\gamma=\sum a_\nu$, $\nu<\gamma$, $n(\nu)=n_0<n(\gamma)$.

Now let (a_ν) be any basis satisfying 12.2. It follows that:

12.3. If (a_{ν_i}) is any finite set of distinct base elements, then (a) $\prod a_{\nu_i}$ is equal either to θ or to the (necessarily unique) element of least subscript ν_k, according as the sequence is or is not nested; (b) if $\nu_i<\nu$ then $a_\nu-\sum a_{\nu_i}$ is the sum of a finite set of disjoint base elements of subscripts ν_i' such that $n(\nu_i')=n(\nu_k)$; (c) any countable sum of base elements can be expressed as the sum of disjoint base elements. (For $\sum a_{\nu_i}=\sum(a_{\nu_i}-\sum_{j<i}a_{\nu_j}a_{\nu_i})$, where the terms of the right side are disjoint, and from (b) each consists of a finite sum of disjoint elements.)

Further, if (a_ν) and (b_ν) are bases of a and b, respectively, which satisfy 12.2:

12.4. For any finite sequence (ν_i) of which we suppose ν_k the least (a) either both $\prod a_{\nu_i}=\prod b_{\nu_i}=\theta$, or $\prod a_{\nu_i}=a_{\nu_k}$ and $\prod b_{\nu_i}=b_{\nu_k}$; (b) if $\nu>\nu_i$, then either $a_\nu=\sum a_{\nu_i}$ and $b_\nu=\sum b_{\nu_i}$, or there exists a sequence (ν_j'), such that $a_\nu-\sum a_{\nu_i}=\sum a_{\nu_j'}$ and $b_\nu-\sum b_{\nu_i}=\sum b_{\nu_j'}$. (c) For corresponding subsequences (a_{ν_i}), (b_{ν_i}) of (a_ν) and (b_ν) there exists a sequence (ν_j') of subscripts such that $(a_{\nu_j'})$, $(b_{\nu_j'})$ are disjoint and $\sum a_{\nu_i}=\sum a_{\nu_j'}$, $\sum b_{\nu_i}=\sum b_{\nu_j'}$.

12.5. Definition. *Any (finite) union of base elements of any a of M is called a finite σ-element of the basis (a_ν). The σ-element $\sum a_{\nu_i}$ will be denoted by* $s_{[\nu_i]}(a)$.

From the previous results it follows that the σ-elements of any basis (a_ν) of an element a have the properties listed below:

12.6. *All finite products, all countable sums, and the relative complement of a finite σ-element with respect to any σ-element are all either σ-elements or equal to θ. All σ-elements can be expressed as the sum of a sequence of disjoint base elements.*

12.7. *Any element $b<a$ is the product of a nested sequence of σ-elements $s_k(a)$ such that* $\lim |s_k(a)|=|b|$.

For from the definition of a basis, for any sequence of positive numbers $\epsilon_n\to 0$, there exists a sequence of σ-elements $s_n(a)>b$, and such that $|s_n-b|<\epsilon_n$. But from 12.6, the elements $\prod_{i=1}^k s_n$ are σ-elements and satisfy the required conditions.

12.8. *If* (a_ν), *and* (a'_ν) *are bases of* a *and* a', *respectively, and if the corresponding* σ*-elements* $s_\alpha(a) \leqq \sum a_{\nu_i(\alpha)}$ *and* $s_\alpha(a') \equiv \sum a_{\nu_i(\alpha)}$ *are denoted by* s_α *and* s'_α, *then, for any sets of corresponding* σ*-elements* s_i *and* s'_i *either both members of any one of the pairs* $\prod_1^k s_i$ *and* $\prod_1^k s'_i$ *(where* k *is any finite number),* $\sum s_i$ *and* $\sum s'_i$, *and the pair of symmetric differences* $s_1 - s_2$ *and* $s'_1 - s'_2$, *are equal to* θ, *or in each case there exists a pair of corresponding* σ*-elements* s_{α_1} *and* s'_{α_1}, *or* s_{α_2} *and* s'_{α_2}, *or* s_{α_3} *and* s'_{α_3}, *such that* $\prod_1^k s_i = s_{\alpha_1}$ *and* $\prod_1^k s'_i = s'_{\alpha_1}$, *or* $\sum s_i = s_{\alpha_2}$ *and* $\sum s'_i = s'_{\alpha_2}$, *or* $s_1 - s_2 = s_{\alpha_3}$ *and* $s'_1 - s'_2 = s_{\alpha_3}$.

13. **Infinite products. Regular bases.** We prove first

13.1. LEMMA. *If* (a_{ν_i}) *is any infinite sequence of base elements with distinct subscripts, then* $\prod a_{\nu_i} = \theta$.

Proof. We may assume that the sequence can be arranged so that $\nu_i < \nu_{i-1}$, and so that $a_{\nu_i} < a_{\nu_{i-1}}$, for otherwise from (a) of 12.3 the conclusion that $\prod a_{\nu_i} = \theta$ already holds. If under this restriction $\prod a_{\nu_i} = a' \neq \theta$, there exists a ν_0 such that $|a_{\nu_0} a'| \neq 0$. Then since for all i, $a_{\nu_0} a_{\nu_i} \neq \theta$, it follows for all but the finite set of the a_{ν_i} for which $n(\nu_i) < n(\nu_0)$, that $a_{\nu_0} < a_{\nu_i}$, and so further that $a_{\nu_0} > a'$. By iii of (ζ), since $a' \neq \theta$, there exists an $a'' < a'$ such that $\theta < a'' < a'$. But every σ-element s covering a'' must also contain a', and hence $|s - a''| > |s - a'| > 0$, which contradicts iv of (ζ). Therefore $a' = \theta$.

13.2. LEMMA. *If* (a_ν) *is a basis of* a *satisfying* 12.2, *for any finite positive* $\xi < |a|$ *there exists a* σ*-element* s *such that* $|s| = \xi$. *More specifically, to any such* ξ *we can make correspond a unique* σ*-element* s_ξ *such that* $|s_\xi| = \xi$, *and if* $0 < \xi_1 < \xi_2 \leqq |a|$, $s_{\xi_1} < s_{\xi_2}$. *Hence further each* σ*-element corresponds to at most one number* ξ.

Proof. Since, for any n, $\sum a_\nu = a$, where $\nu \in \Gamma_n$, if we order the a_ν of Γ_n according to the first differences into a sequence $a_{n,k}$ and if we let $s_{n,k} = \sum_1^n a_{n,i}$ then for each $\xi > 0$ there corresponds a $k = k(\xi, n)$ such that

$$(1) \qquad |s_{n,k}| \leqq \xi \leqq |s_{n,k+1}|,$$

and further

$$(2) \qquad s_{n,k_n} < s_{n+1,k_{n+1}} < s_{n+1,k_n+1} < s_{n,k_n+1}.$$

From (1) and the fact that the elements $a_{n,k}$ are nested, we have

$$(3) \qquad \xi - |\lim s_{n,k}| = \lim |a_{n,k}| = 0.$$

Therefore $|\sum s_{n,k}| = \xi$ and since if $\xi_1 < \xi_2$ then for each n $k_{n,\xi_1} < k_{n,\xi_2}$, we also have $\sum s_{n,k(\xi_1)} < \sum s_{n,k(\xi_2)}$, and we can define $s_\xi = \sum s_{n,k_{n,\xi}}$.

13.21. LEMMA. *Every element* a *such that* $|a| = \infty$, *is the sum of a set of disjoint elements* a_n *such that* $|a_n| = 1$ *for each* n.

Proof. Every such element a is the sum of a set of elements a'_n such that

$|a_n'| < \infty$, and which we may suppose disjoint and different from θ. Let k_i be a monotonic set of integers such that $|a_{k_i}| \leqq i < |a_{k_{i+1}}|$, and let us define a nested sequence of elements a_i^* as follows: If the lower inequality holds, let $a_i^* = a_k'$; if $i > |a_{k_i}'|$, let $a_i^* = a_{k+a''_i}$, where $a_i'' \leqq a'_{k_{i+1}} - a'k_i$, and $|a_i''| = i - |a_{k_i}'|$. In either case, $|a_i^*| = i$, and $a_{k_i} \leqq a_i^* < a_{k_{i+1}}$, whence lim $a_i^* = a$. We can therefore define the required elements a_n as a_1^*, $a_n^* - a^*_{n-1}$.

13.3. Definition. *Let ν be any number of the form $k + m/2^n$, where k is a positive integer, $0 \leqq m < 2^n$, and $m/2^n$ is expanded dyadically into a sequence of $n(\nu)$ terms. Then a basis (a_ν) of an element $a \neq \theta$ is called regular when $|a| < \infty$ if $0 \leqq \nu < 1$ and $|a_\nu| = |a|/2^{n(\nu)}$, and when $|a| = \infty$ if $0 \leqq \nu < \infty$ and $|a_\nu| = 1/2^{n(\nu)}$.*

13.4. Theorem. *Every $a \neq \theta$ has a regular basis.*

Proof. Suppose $|a| < \infty$, and let (a_ν') be any basis satisfying 12.2. Then there exist σ-elements $s_{n,m}$ for $n = 1, 2, \cdots$, and $m = 1, 2, \cdots, 2^n$, such that $|s_{n,m}| = m|a|/2^n$, and such that $m/2^n > m'/2^{n'}$ implies that $s_{n,m} > s_{n',m'}$. Then if ν is the dyadic expansion of $m/2^n$, $m < 2^n$, then we define a_ν as $s_{n,m+1} - s_{n,m}$, and we make the convention that $a_0 = 0$. If $s'_{n,m} = \sum a_{\nu_i}'$, where ν_i are the dyadic expansions of the numbers $m'/2^n$, $m' \leqq m < 2^n$, then $a_\nu' = s'_{n,m+1} - s'_{n,m} = \sum a_{\nu_j}$, where $|s'_{n,m}| \leqq \nu_j \leqq |s_{n,m+1}|$. Therefore from 12.1, (a_ν) is a basis, which can easily be seen to be regular. If $|a| = \infty$, let (a_k) be a sequence satisfying 13.21, and combine their bases for a basis for a.

13.5. *If $|a| = |a'| \neq 0$, and if (a_ν), (a_ν') are regular bases of a and a', respectively, and if as previously we denote the corresponding σ-elements $s_\alpha(a) = \sum a_{\nu_i(\alpha)}$ and $s_\alpha' = \sum a'_{\nu_i(\alpha)}$ by s_α and s_α' respectively, then for any set of corresponding pairs it follows from* 12.8 *and from the fact that $|a_\nu| = |a_\nu'|$ that $|s_\alpha| = |s_\alpha'|$; $|s_1 - s_2| = |s_1' = s_2'|$, where the differences are symmetric; $|\sum s_i| = |\sum s_i'|$; and $|\prod_i^k s_i| = |\prod_i^k s_i'|$.*

14. **The isomorphism ϕ.** We are now in a position to establish the theorem referred to in §12, namely, that

14.1. Theorem. *If $|a| = |a'|$, then there exists a σ-isomorphism ϕ of $I(a)$ (see* 1.1*) onto $I(a')$ such that for any $b \leqq a$, $|\phi(b)| = |b|$.*

Proof. The theorem is trivial if $a = a' = \theta$. Otherwise, let (a_ν), (a_ν'), be regular bases of a and a' respectively. Then by 12.8 for any $b \leqq a$ there exists a nested sequence of σ-elements $s_k > b$ such that lim $|s_k - b| = 0$. We define for any $b < a$.

$$\phi(b) \equiv \phi\left(\prod s_k\right) \equiv \prod \phi(s'_k). \tag{1}$$

Obviously, ϕ assigns at least one image to every element of $I(a)$ and at least one counterimage to every element of $I(a')$. Further from 13.5 for each k, $|s_k| = |s_k'|$, whence for any image $\phi(b)$

$$|\phi(b)| = \lim |s_k'| = \lim |s_k| = |b|, \tag{2}$$

and *the functional* $|a|$ *is invariant under* ϕ. Also, if $b_1 \equiv \prod s_{k,1}$ and $b_2 \equiv \prod s_{k,2}$ are any two elements of $I(a)$, then for each k, $|s_{k,1}-s_{k,2}| = |s'_{k,1}-s'_{k,2}|$ where the differences are symmetric. Passing to the limit with respect to k,

$$(3) \qquad b_1 = b_2 \text{ if and only if } \phi(b_1) = \phi(b_2),$$

that is, *the isomorphism* $\phi(b)$ *is univocal.*

Now suppose $b_1 \equiv \prod s_{k,1} \leqq b_2 \equiv \prod s_{k,2}$. Then from 12.8, the elements $s_{k,3} = s_{k,1}s_{k,2}$ form a nested sequence of σ-elements such that $b_1 \leqq s_{k,3} \leqq s_{k,2}$. Again from 13.5 for each k, $s'_{k,3} = s'_{k,1}s'_{k,2}$ so

$$(4) \qquad \phi(b_1) = \prod s'_{k,3} \leqq \prod s'_{k,2} = \phi(b_2).$$

Applying this same argument with reference to ϕ^{-1}, we see that

$$(5) \qquad b_1 \leqq b_2 \text{ if and only if } \phi(b_1) \leqq \phi(b_2).$$

Let (a_r) be a sequence of elements of $I(a)$, and for each r let $(s_{k,r})$ be a nested sequence of σ-elements such that $a_r = \prod_k s_{k,r}$ and such that $|s_{k,r}-a_r| < 1/2^{k+r}$. Then for each k, $\sum_r s_{k,r}$ is a σ-element and

$$(6) \qquad \sum a_r \leqq \sum_r s_{k,r} \leqq \sum_r s_{k-1,r},$$

$$(7) \qquad \begin{aligned} \left| \sum_r s_{k,r} - \sum a_r \right| &= \left| \sum_r \left[s_{k,r} - \left(\sum_r a_r \right) s_{k,r} \right] \right| \\ &\leqq \sum_r | s_{k,r} - a_r | \leqq 1/2^k, \end{aligned}$$

and so

$$(8) \qquad a_r = \prod_k \left(\sum_r s_{k,r} \right).$$

From (5), for each r, $\phi(a_r) \leqq \phi(\sum a_r)$, and so further

$$(9) \qquad \sum \phi(a_r) \leqq \phi(\sum a_r).$$

But for each k,

$$(10) \qquad \begin{aligned} \left| \sum_r s'_{k,r} - \sum \phi(a_r) \right| &\leqq | s'_{k,r} - \phi(a_r) | = \sum [| s'_{k,r} | - | \phi(a_r) |] \\ &= \sum (| s_{k,r} | - | a_r |) \leqq 1/2^k, \end{aligned}$$

and passing to the limit with respect to k gives

$$(11) \qquad \sum \phi(a_r) = \phi(\sum a_r).$$

From 12.6 for each finite r the elements $\prod_0^r s_{k,i}$ form a nested sequence of σ-elements such that $\prod_0^\infty a_i \leqq \prod_0^r a_i \leqq \prod_0^r s_{k,i}$. If for some i $|a_i| < \infty$, and if (r_i) is an increasing sequence of indices such that $|\prod_0^{r_i} a_r| - |\prod_0^\infty a_r| < 1/2i$,

$$(12)\qquad \left|\prod_0^{r_i} s_{k,r} - \prod_0^{\infty} a_{k,r}\right| \leqq \left|\prod_0^{r_i} s_{k,r} - \prod_0^{r_i} a_r\right| + \left|\prod_0^{r_i} a_r - \prod_0^{\infty} a_r\right|.$$

But since $s_{k,r} \leqq \prod_0^{r_i} a_r + \sum_0^{r_i}(s_{k,r} - a_r)$, (12) yields

$$(13)\qquad \left|\prod_0^{r_i} s_{k,r} - \prod_0^{\infty} a_r\right| < 1/2^{k+i},$$

whence

$$(14)\qquad \prod_0^{\infty} a_r = \prod_{k=0}^{\infty}\left(\prod_{r=0}^{r_k} s_{k,r}\right).$$

From (5), for each r, $\phi(\prod_r a_r) \leqq \phi(a_r)$, and so

$$(15)\qquad \phi(\prod a_r) \leqq \prod \phi(a_r).$$

On the other hand,

$$(16)\qquad \sum \phi(a_r) = \prod_{r=0}^{\infty}\prod_{k=0}^{\infty} s'_{k,r} = \prod_{k=0}^{\infty}\prod_{r=0}^{\infty} s'_{k,r} \leqq \prod_{k=0}^{\infty}\prod_{r=0}^{r_k} s'_{k,r} = \phi(\prod a_r),$$

which with (15) gives

$$(17)\qquad \prod \phi(a_r) = \phi(\prod a_r),$$

if for some i, $|a_i| < \infty$. If for all r, $|a_r| = \infty$, let $(a_{1,j})$ be a sequence of elements such that $\sum a_{1,j} = a_1$ and $|a_{1,j}| < \infty$ for all j's. Then from (11) and (17),

$$(18)\qquad \phi(\prod a_r) = \phi\left[\sum_j\left(\prod_r a_{1,j}a_r\right)\right] = \sum_j\left[\prod_r \phi(a_{1,j})\phi(a_r)\right] = \prod \phi(a_r),$$

and this holds for any sequence (a_r).

It is now verified by statements (2), (3), (11), and (18) that ϕ satisfies the conditions required in Theorem 14.1, the proof of which is now completed. We have as an immediate corollary to 14.1 that

14.11. The class Φ of all σ-isomorphisms ϕ between pairs of principal ideals of M is non-void and satisfies (β) of §1.

15. Application to Part II. Examples.

15.1. If now $\boldsymbol{m}$ is a field of sets upon which a measure function $|\boldsymbol{a}|$ is defined, and if $\boldsymbol{n}$ is the ideal of sets for which $|\boldsymbol{a}| = 0$, then we define a functional $|a|$ on the residue class ring $M = \boldsymbol{m}/\boldsymbol{n}$ as the common value of $|\boldsymbol{a}|$ for all the sets $\boldsymbol{a} \in a$. If $|a|$ is such that (ζ) holds, then by Theorem 15.1 we have a measuring system $(\boldsymbol{m}, \boldsymbol{n}, \Phi)$ (§6). If as before we denote the measure in the sense of Part II of any $a \in M$ by $\mathfrak{a}$, then $\mathfrak{a} = \mathfrak{b}$ if and only if $|a| = |b|$, $\mathfrak{a} = 0$ if and only if $a = \theta$, that is, if and only if $|a| = 0$; and $\mathfrak{a} = \sum \mathfrak{a}_n$ if and only if $|a| = \sum |a_n|$. Further from 13.2, $\mathfrak{a}$ is bounded in the sense of Part II if, and from 15.21 only if, $|a| < \infty$. From the fact that $a < b$ if and only if $|a| < |b|$,

it is clear that the system (m, n, Φ) satisfies (δ) (§9) and (ϵ) (§10). Therefore the correspondence $f(\mathbf{a}) \equiv f(a) = |\mathbf{a}|$ is an isomorphism of the type described in Theorem 10.1 of the set $\mathfrak{M}$ of measure values onto a closed interval. It follows from 10.1 that the ratios $|a|/|b|$ are uniquely determined by the system (m, n, Φ) as the numbers $\mu(a, b)$ of the lemma to 10.1, and so that $|\mathbf{a}|$ is determined up to a multiplicative constant.

15.2. Examples. Let $\mathbf{s}$ be any locally compact metric spaces and let $|\mathbf{a}|$ be a positive functional on the open sets with compact closures such that $|\mathbf{a}| = 0$ only on the empty set, and such that with any set of concentric spheres $\mathbf{a}_\delta$ of radius ϵ, $\lim |\mathbf{a}_\epsilon| = 0$. Then the measure generated in the usual fashion from $|\mathbf{a}|$ satisfies conditions (ζ) and so on can be obtained as in Part II. The Haar measure on a locally compact group is also covered by Part II, since the decomposition theorem of Banach and Tarski can be extended([3]) to this case; if the fundamental space is not a group, but metric and locally compact([4]), we need make only a trivial examination of the case in which points have positive measure.

15.3. The theorem of Banach and Tarski mentioned in the Introduction can be generalized under conditions of 15.1 in the following sense: Let (s_ν) be a regular basis of s (13.3). An isomorphism ϕ of Φ is called a *congruence* if ϕ is defined between two base elements, say s_ν and $s_{\nu'}$, where obviously $n(\nu) = n(\nu')$, and satisfies the following condition:

Let ν, ν', be the dyadic expansions of $m/2^n$, $m'/2^n$, respectively. The elements $s_\gamma < s_\nu$ $[s_{\gamma'} < s_\nu]$ are those such that $m/2^n \leqq \gamma < m + 1/2^n$, $[m'/2^n \leqq \gamma' < m' + 1/2^n]$. We define $\phi(s_\gamma)$ for all such γ as $s_{\gamma'}$, where $\gamma' = m'/2^n + (\gamma - m/2^n)$. Obviously, ϕ is completely determined for any $s' < s$ by the above relation. Then under conditions (ζ) and with this definition for congruence, a necessary and sufficient condition that two elements a and a' be of equal measure is that there exist sequences (a_n), (a_n') of disjoint elements such that $a = \sum a_n$, $a' = \sum a_n'$ and a_n and a_n' are congruent for each n.

The sufficiency is obvious. The necessity will be proved for the case for which $|a| = |a'| < \infty$ and both a and a' are different from s, from which the general case follows at once. The required proof then reduces to that of the following statements: a. If $|a| = |a'| < \infty$, then each element a or a' is the limit of a nested sequence of σ-elements of finite measure, of which corresponding elements have equal measure. b. Each of two σ-elements of equal measure is the sum of a family of disjoint base elements, corresponding elements having equal measure and so being congruent. For if a and b hold, and if we let the class of all sums of congruences, which class can be seen to satisfy (β), be Φ, then we can apply Theorem 4.11 (which states that the respective limits of

([3]) See Footnote 1.

([4]) As in the "Note on Haar's measure," by S. Banach, in Saks' *Theory of the Integral*, Appendix II.

monotonic sequences of bounded equivalent elements are equivalent) and so get the desired theorem.

To prove a, we consider nested sequences of σ-elements $s_k > a$, and $s_k' > a'$, such that for some $\epsilon > 0$, $|s| > \epsilon \geqq |s_k - a| \geqq 2|s_{k+1} - a|$, and $|s_k' - a'| < |s_k - a|$. It follows that $(|s_k| - |s_k'|) < (|s_{k-1}| - |s_{k-1}'|)$. But for each p, the elements (s_ν) for which $n(\nu) = p$ can be ordered into a sequence $(s_{n,p})$. Then for each k and p, there is an integer $i = i_{k,p}$ for which $|\sum_{j=0}^{i-1} s_{j,p} + s_k'| \leqq |s_k| - |s_k'| < |\sum_{j=0}^{i} s_{j,p} + s_k'|$, and the proof proceeds as in 13.2. Relation b is obvious in the case of two finite sums of base elements, for each can be expressed as the sum of elements s_γ such that $n(\nu) = n^*$, where n^* is the maximum value of $n(\nu)$ occurring in the given summands. Otherwise we write both summands as infinite sums. But two infinite sums can, as a result of 13.21 and the regularity of the basis (s_ν), be expressed as sums of infinite disjoint sets of finite sums, of which corresponding finite sums have equal measure. Hence our assertion is proved.

In the case of euclidean r-space $(\xi_1, \cdots, \xi_r)$ we may consider our class Φ as sums of rigid motions and the regular basis as the family of cubes $[m/2^n \leqq \xi_i < (m+1)/2^n]$, $i = 1, \cdots, r$. Our result then reduces to the theorem of Banach and Tarski from the fact that every open set is the union of a set of disjoint half-open cubes, and that every measurable set is G_δ, modulo the null sets.

It can be deduced from the results of Part III that under the conditions of 15.1 that $M = m/n$ and the totality Φ of measure-preserving transformations on M are respectively isomorphic to the field M' of measurable sets modulo the null sets in euclidean space, and to the totality Φ' of measure-preserving transformations on M'.

Bryn Mawr College,
Bryn Mawr, Pa.

THE CONFORMAL THEORY OF CURVES

BY

AARON FIALKOW

TABLE OF CONTENTS

1. **Introduction.** Classical differential geometry is the metric theory of euclidean 3-space R_3. Its generalization, Riemannian geometry, is the metric theory of an n-dimensional Riemannian manifold(1) V_n. On the whole, the development of these geometries has proceeded in two main directions. Naturally these two directions are not mutually exclusive; occasionally they overlap in the common development of some subject.

One approach is the study of the metric transformations of the manifolds as a whole upon each other. This is the intrinsic theory of the space. In classical geometry, this point of view yields rather meager results since the intrinsic theory of R_3 is almost synonymous with the discovery of the complete group of motions in R_3. In Riemannian geometry, the intrinsic theory has considerably greater significance. The discovery of the process of covariant differentiation with respect to the first fundamental form of V_n and of the Riemann curvature tensor of V_n are important milestones in the development of this theory. This approach reaches its culmination in the fundamental theorem which states the conditions under which two Riemann spaces are isometric.

The other approach is the study of curves, surfaces and other subspaces and configurations in the enveloping R_3 or V_n and their behavior when the

Presented to the Society October 28, 1939; received by the editors April 11, 1940, and, in revised form, March 27, 1941.

(1) We denote an n-dimensional Riemann space, Einstein space, euclidean space and a space of constant curvature by V_n, E_n, R_n and S_n respectively.

enveloping manifold undergoes any metric transformation. In its most common development today, this study is based upon the process of covariant differentiation. As is well known, by repeated use of this type of differentiation, a system of Frenet equations of the subspace is obtained. These equations involve a number of metric geometric objects([2]): the $(n-1)$ curvatures and arc length for curves, the coefficients of the first and second fundamental forms for hypersurfaces, and so on. These geometric objects constitute the foundation upon which the detailed geometry of curves, surfaces and subspaces is built.

Classical differential geometry concerns itself almost exclusively with this second approach and a very considerable portion of Riemannian geometry has also evolved in this direction. The development of conformal Riemannian geometry, however, presents a different picture. Here the main emphasis has been upon the intrinsic conformal theory of the manifolds; that is, the investigation of the conformal transformations of Riemann spaces as a whole upon each other. This point of view is maintained in the early papers of Weyl([3]) and Schouten([4]) on conformal Riemannian geometry which mark the modern beginning of that subject. The fundamental conformal curvature tensor is discovered in these papers and is used in order to obtain a complete characterization of conformally euclidean Riemann spaces. These results are a continuation of classical theorems such as the theorem of Liouville on the conformal transformations of R_3 on itself.

The central problem of the intrinsic theory is the question of the conformal equivalence of Riemann spaces V_n. In order to effect a solution of this problem, T. Y. Thomas has considered the conformal tensor $g_{ij}/g^{1/n}$ where g_{ij} is the metric tensor of V_n and g is the determinant $|g_{ij}|$. This tensor remains invariant under conformal transformations of the metric tensor of V_n. The Christoffel symbols([5]) formed with respect to this tensor (called the conformal parameters) have a complicated law of transformation under coordinate transformations and one cannot define a simple covariant derivative of tensors by means of these parameters. However, by formal methods based upon

---

([2]) By a geometric object we mean an abstract object having a unique set of components, depending on the coordinates and their differentials to a specified order, in any coordinate neighborhood of the manifold. Hence the law of transformation of the components under coordinate changes must be transitive.

([3]) H. Weyl, *Reine Infinitesimalgeometrie*, Mathematische Zeitschrift, vol. 2 (1918), pp. 384–411.

([4]) J. A. Schouten, *Über die konforme Abbildung n-dimensionaler Mannigfaltigkeiten mit quadratischer Massbestimmung auf eine Mannigfaltigkeit mit euklidischer Massbestimmung*, Mathematische Zeitschrift, vol. 11 (1921), pp. 58–88.

([5]) These Christoffel symbols were first defined by J. M. Thomas in another way. Cf. J. M. Thomas, *Conformal invariants*, Proceedings of the National Academy of Sciences, vol. 12 (1926), pp. 389–393.

the conformal parameters, it is possible to obtain a solution to the conformal equivalence problem for Riemann spaces(6).

The investigations of conformal Riemannian geometry by Cartan(7) and Schouten(8) affords another method for the development of this subject. This method depends upon the introduction of $(n+2)$ homogeneous coordinates (the generalization of tetracyclic and pentaspherical coordinates) into the local euclidean space R_n of the V_n. Still another path for the study of intrinsic conformal geometry is indicated by the recent results of Schouten and Haantjes(9) which suggest a projective treatment of conformal geometry.

Thus there exists a variety of general methods for the development of the intrinsic conformal theory of Riemann spaces. While this formal intrinsic theory is complete, the conformal theory of configurations in a Riemann space has been largely neglected. This fact is all the more remarkable when one considers that such a theory would always have real significance whereas this is rarely the case for the corresponding metric theory of configurations in a general V_n. To illustrate this point, we note that while a curve in a general V_n has $(n-1)$ curvatures which are metric invariants, these invariants are not very meaningful if the V_n does not admit any metric transformations other than the identity (as is usually the case). This state of affairs is never encountered in the conformal theory of configurations since every V_n always admits an infinite number of conformal mappings on conformally equivalent Riemann spaces.

One of the earliest results belonging to the conformal theory of configurations in V_n is the theorem which states that the lines of curvature of a hypersurface of V_n remain invariant under conformal transformations of V_n, first proved for a general V_n by Schouten and Struik(10). They also proved a considerable number of similar results, some of which are not purely conformal theorems since they depend upon metric properties of the configuration and upon the particular conformal transformation to which the V_n is subjected.

(6) T. Y. Thomas, *The Differential Invariants of Generalized Spaces*, 1934, chap. 4. O. Veblen, *Formalism for conformal geometry*, Proceedings of the National Academy of Sciences, vol. 21 (1935), pp. 168–173.

(7) E. Cartan, *Les espaces à connexion conforme*, Annales de la Société Polonaise de Mathématique, vol. 2 (1923), pp. 171–221.

(8) J. A. Schouten, *On the place of conformal and projective geometry in the theory of linear displacements*, Proceedings, K. Akademie van Wetenschappen, Amsterdam, vol. 27 (1924), pp. 407–424.

(9) J. A. Schouten and J. Haantjes, *Beiträge zur allgemeinen (gekrümmten) konformen Differentialgeometrie*. I, II, Mathematische Annalen, vol. 112 (1936), pp. 594–629; vol. 113 (1936), pp. 568–583.

(10) J. A. Schouten and D. J. Struik, *Un théorème sur la transformation conforme dans la géométrie différentielle à n dimensions*, Comptes Rendus de l'Académie des Sciences, vol. 176 (1923), pp. 1597–1600. Also cf. J. A. Schouten and D. J. Struik, *Einführung in die neueren Methoden der Differentialgeometrie*, vol. 2, 1938, pp. 199–215 for this topic as well as a general discussion of conformal Riemannian geometry.

The recent investigations of Sasaki(11), Modesitt(12), and the author(13) are of this general character. The papers of Kasner(14), Lipke(15), Schouten(16), and the writer(17) on conformal geodesics (natural families of curves) also belong in this category.

A number of investigators have used the formal methods which were devised to obtain a solution of the equivalence problem in the conformal intrinsic theory, in order to develop a conformal theory of curves and other subspaces. Among these developments are the results(18) of Sasaki(19) and Yano(20). The formal apparatus used in these papers is necessarily quite complicated because their methods follow those used in the intrinsic theory. The various derivatives which have been devised for the development of the intrinsic theory have a strongly formal character and their structure is more complicated than that of ordinary covariant differentiation.

But while this formal apparatus may be inevitable in the case of the intrinsic theory, it is not essential for the development of the conformal theory of a subspace. For the subspace introduces additional structure into the Riemann space V_n by means of which we find a relative conformal scalar(21) at points of the subspace. By means of this relative conformal scalar, it is possible to define a new simple type of differentiation (with respect to the subspace) which plays a role analogous to ordinary covariant differentiation in metric Riemannian geometry. This differentiation process enjoys all the usual properties of covariant differentiation as well as a number of others which give it its distinctive conformal character.

---

(11) S. Sasaki, *Some theorems on conformal transformations of Riemannian spaces*, Proceedings of the Physico-Mathematical Society of Japan, (3), vol. 18 (1936), pp. 572–578.

(12) V. Modesitt, *Some singular properties of conformal transformations between Riemann spaces*, American Journal of Mathematics, vol. 60 (1938), pp. 325–336.

(13) A. Fialkow, *Conformal transformations and the subspaces of a Riemann space*, Bulletin of the American Mathematical Society, abstract 43-9-328.

(14) E. Kasner, *Natural families of trajectories: conservative fields of force*, these Transactions, vol. 10 (1909), pp. 201–219.

(15) J. Lipke, *Natural families of curves in a general curved space of n dimensions*, these Transactions, vol. 13 (1912), pp. 77–95.

(16) J. A. Schouten, *Über die Umkehrung eines Satzes von Lipschitz*, Nieuw Archief voor Wiskunde, vol. 15 (1928), pp. 97–102.

(17) A. Fialkow, *Conformal geodesics*, these Transactions, vol. 45 (1939), pp. 443–473.

(18) Possibly the work of Hlavatý also belongs in this category. These papers are not accessible to the writer. Cf. V. Hlavatý, *Zur Konformgeometrie* III, Proceedings, K. Akademie van Wetenschappen, Amsterdam, vol. 38 (1935), pp. 1006–1011.

(19) S. Sasaki, *On the theory of curves in a curved conformal space*, Science Reports of the Imperial University of Tokyo (1), vol. 27 (1939), pp. 392–409; *On the theory of surfaces in a curved conformal space*, ibid., vol. 28 (1940), pp. 261–285; *Geometry of the conformal connexion*, ibid., vol. 29 (1940), pp. 219–267.

(20) K. Yano, *Sur la théorie des espaces à connexion conforme*, Journal of the Faculty of Science, Imperial University of Tokyo, vol. 4 (1939), pp. 40–57.

(21) This term is defined in §2.

By "conformal differentiation," we arrive at a sequence of normal vector spaces and fundamental forms for the subspace which are unchanged by conformal transformations of V_n. These "conformal fundamental forms" constitute the foundation upon which a detailed conformal geometry of subspaces may be built. Formally, this entire theory is considerably simpler than the previous investigations of conformal Riemannian geometry, its technical aspects being no more involved than those of the ordinary metric geometry of Riemann spaces. While we are concerned with the same general subject as that dealt with by Sasaki and Yano, there is no actual overlapping either of results or of methods. We note, however, as is shown in §15, that our results may be used to develop a conformal theory of curves based upon the conformal tensor $g_{ij}/g^{1/n}$ which is formally analogous to the investigations mentioned above.

In the present paper, we develop the foundations of the conformal theory of curves, reserving the treatment of other subspaces for later publication([22]). We note that this separate treatment is not prompted by pedagogic reasons alone, but is a natural separation. For in our development of the conformal geometry of a subspace of V_n, two mutually exclusive cases arise which must be treated separately: (1) curves and (2) subspaces whose dimensionality exceeds one.

It is well known that there is a metric (congruence) theory of curves in the plane but no conformal theory. That an analytic curve can have no conformal properties follows from the theorem: *Every analytic curve in the plane is conformally equivalent to a straight line.* It is the object of this paper to show that a conformal theory of curves *does* exist in any Riemann space whose dimensionality exceeds 2 and to develop this theory. Accordingly, we study those properties of a curve which remain unchanged when the enveloping Riemann space V_n of dimensionality $n>2$ undergoes any conformal mapping, not necessarily on itself.

The principal tool is a new kind of tensor differentiation which has conformal meaning—"the conformal derivative." By systematic use of the conformal derivative we derive the conformal analogues of the ordinary (metric) Frenet equations. We find $n-1$ differential "conformal curvatures" $J_1, J_2, \cdots, J_{n-1}$ and an integral "conformal arc length" S which are unchanged by any conformal transformation of the Riemann space. This means that *if* $V_n \leftrightarrow \overline{V}_n$, $C \leftrightarrow \overline{C}$ *by a conformal map, then the* J's *are the same functions of* S *for* C *and* $\overline{C}$.

The converse holds in spaces which are conformal to a euclidean space.

---

(22) Some of the principal results in the curve theory are stated without proof in a previous note having the same title as the present paper which appeared in the Proceedings of the National Academy of Sciences, vol. 26 (1940), pp. 437–439. Corresponding results in the conformal theory of a subspace appear in two abstracts in the Bulletin of the American Mathematical Society, abstracts 46-11-487 and 47-3-156.

In this case, we have the fundamental conformal equivalence theorem: *If V_n and $\overline{V}_n$ are conformal to a euclidean space and the J's for both C and $\overline{C}$ are the same functions of S, then a conformal mapping exists for which $V_n \leftrightarrow \overline{V}_n$, $C \leftrightarrow \overline{C}$.* This is the conformal analogue of the metric congruence theorem which holds in a euclidean space and in a space of constant curvature.

We also prove the existence theorem: *In any V_n, a curve exists for which the J's are preassigned continuous functions of S. This curve is uniquely determined by a set of initial conditions which is found explicitly.*

The conformal curvatures of a curve C in V_n have rather simple geometric properties if V_n is conformal to an Einstein space or, more particularly, to a euclidean space. Thus *if V_n is conformally euclidean then $J_\alpha \equiv 0$ $(1 \leqq \alpha \leqq n-2)$ if and only if C is conformally equivalent to a curve in a euclidean n-space whose $(\alpha+1)$st metric curvature vanishes. Another example: If V_n is conformal to a euclidean space then the $n-1$ conformal curvatures of a curve and their derivatives with respect to the conformal length constitute a complete set of conformal differential invariants of the curve.*

If $n=2$, the results of this paper apply if the conformal transformations are restricted to mappings applied to spaces of constant curvature which are similar to and include the inversive transformations of the plane.

While the results of this paper bear a close analogy to those which hold in the metric theory, in some cases the proofs are markedly different. Thus, the first of the "conformal Frenet equations" is not obtained, as in the classic case, by differentiating the unit tangent vector. For it will be seen later that the conformal derivative of the unit tangent always vanishes identically. As another important point of difference, we note that only $n-2$ of the conformal curvatures occur as coefficients in the conformal Frenet equations. The $(n-1)$st conformal curvature, while as essential as the other curvatures, is found in an entirely different way and does not have the same properties as the others.

These essentially novel features which distinguish the conformal geometry of curves from the metric geometry are also present in an analogous form in the corresponding theory for any subspace. For example, a hypersurface has three "conformal fundamental forms" instead of the anticipated two forms and four sets of integrability conditions instead of the classic Gauss-Codazzi equations. Furthermore, the conformal behavior of subspaces whose dimensionality is at least 4 is typical, while the cases of dimension number 3, 2, and 1 respectively are increasingly degenerate. There is no corresponding analogue in the metric theory.

As an important special case, this theory obviously includes the "natural geometry" of curves in euclidean n-space under the continuous group of conformal mappings of the euclidean space upon itself. The transformations of this group are the products of inversions with respect to a hypersphere, motions and transformations of similitude (Liouville's theorem). This means

that our results constitute the inversive theory of curves when applied to this continuous group of transformations of a euclidean space. In particular, *the curves along which all the conformal curvatures are constants are the paths of the inversive group.*

A detailed inversive geometry of plane curves and of curves and surfaces in R_3 has been developed by Thomsen, Blaschke and Takasu in their books on conformal differential geometry([23]). Their investigations constitute a theory of curves and surfaces in R_2 and R_3 which is complete in its essential parts and anticipates many of our results for this important but special case. However, their methods depend upon the systematic use of tetracyclic and pentaspherical coordinates and therefore differ completely from the methods which are employed here. The inversive theory of plane curves has also been developed by a number of other writers using still different methods.

We note that the subject of this paper is also somewhat connected with the "natural geometry" of a curve associated with any group of transformations of the plane into itself. This theory was originated by Pick([24]) and has subsequently been developed by Kowalewski([25]) and his students.

2. **Riemann spaces conformal to V_n, conformal tensors.** Let V_n be a real Riemann space whose coordinate manifold is of class([26]) C^m and whose real metric tensor, defined over the manifold, is positive definite([27]) and of class C^{m-1} with $m \geqq 1$. Briefly, we say V_n is a Riemann space of class([28]) C^m.

([23]) W. Blaschke and G. Thomsen, *Vorlesungen über Differentialgeometrie*, vol. 3, 1929; T. Takasu, *Differentialgeometrien in den Kugelräumen*, vol. 1, 1938.

We are obliged to a referee for these references. Due to our unfamiliarity with tetracyclic and pentaspherical coordinates, it is difficult for us to determine precisely the extent to which duplication of results occurs. In general, these books would appear to contain most of our theorems for curves in R_2 and for curves and surfaces in R_3 under the inversive group. These books also contain other results on the detailed inversive geometry of R_2 and R_3 which lie beyond the scope of our present investigations. These references have been incorporated into the revision of the introduction and we have also included references to a number of papers which have appeared since this paper was first written.

([24]) G. Pick, *Natürliche Geometrie ebener Transformationsgruppen*, Sitzungsberichte der Academie der Wissenschaften, Vienna, vol. 115 (1906), p. 139.

([25]) G. Kowalewski, *Vorlesungen über allgemeine natürliche Geometrie und Liesche Transformationsgruppen*, 1931, chap. 3.

([26]) The definitions of the class of a coordinate manifold and of a Riemann space are based upon the discussion which appears in the paper by T. Y. Thomas, *Recent trends in geometry*, American Mathematical Society Semicentennial Publications, vol. 2 (1938), pp. 98–99, 104. In particular, if the coordinate manifold is of class C^m, then the admissible coordinate systems are related to each other by transformations of class C^m.

([27]) The greater part of the following discussion and of the results of the paper will hold even if the metric tensor is indefinite provided that it is not singular. The only real novelty arises when a vector is a null vector. We shall not consider the indefinite case.

([28]) We shall assume the reality, existence and continuity of whatever functions occur in the proofs. At the outset of the proof of an important theorem we shall simply indicate sufficient conditions for the satisfaction of this assumption in order to avoid frequent interruptions of the discussion for essentially non-geometric matters.

Suppose[29] $\{x^i\}$ are admissible real local coordinates in a coordinate neighborhood of any point of V_n. In each coordinate neighborhood, we write the first fundamental form of V_n as

$$ds^2 = g_{ij}dx^i dx^j. \tag{2.1}$$

Since the results of this paper are local theorems which hold for a sufficiently small neighborhood of a point we shall restrict ourselves to a portion of V_n which is a neighborhood $U(P)$ of a point P coverable by a single coordinate system $\{x^i\}$. We shall refer to $U(P)$ as the Riemann space V_n and use similar language in connection with other Riemann spaces which appear in the paper.

Let[30] $\bar{V}_n$ be a real Riemann space of class C^m whose first fundamental form may be written as

$$d\bar{s}^2 = \bar{g}_{ij}d\bar{x}^i d\bar{x}^j \tag{2.2}$$

where $\{\bar{x}^i\}$ are allowable local coordinates. *Then*[31] $\bar{V}_n$ *is conformal to* V_n *by means of a transformation of class* C^m (*briefly*: $\bar{V}_n$ *is conformal to* V_n) if a one-to-one point transformation T exists between the points P of V_n and the points $\bar{P}$ of $\bar{V}_n$ which may be written (locally) as

$$\bar{x}^i = \bar{x}^i(x^1, x^2, \cdots, x^n), \qquad x^i = x^i(\bar{x}^1, \bar{x}^2, \cdots, \bar{x}^n) \tag{2.3}$$

so that the real functions

$$\bar{x}^i(x^1, x^2, \cdots, x^n), \qquad x^i(\bar{x}^1, \bar{x}^2, \cdots, \bar{x}^n)$$

are of class C^m and

$$d\bar{s} = e^{\sigma} ds \tag{2.4}$$

at corresponding points. It follows that $\sigma(x^i)$ is a real function of class C^{m-1} and that the form (2.2) is positive definite. We refer to $\sigma(x^i)$ as *the conformal mapping function of* V_n *on* $\bar{V}_n$, or briefly, as the *mapping function*. Whenever we say that $\bar{V}_n$ is conformal to V_n, it is to be understood that the conformal transformation is of class C^m.

The transformation T may be written in the simple form

$$\bar{x}^i = x^i \tag{2.5}$$

after a suitable change of coordinates. For if we transform the coordinate

---

(29) Throughout this paper the indices h, i, j, k have the range $1, 2, \cdots, n$. It is to be understood that a tensor equation in which an index is not summed is valid for each value of the index within its range. A covariant or contravariant index which appears twice in an expression is to be summed over the appropriate range.

(30) We denote a Riemann space conformal to V_n by $\bar{V}_n$. Thus $\bar{E}_n$ and $\bar{R}_n$ signify spaces which are conformally equivalent to an Einstein space and a euclidean space respectively. A geometric object in $\bar{V}_n$ corresponding to the geometric object F in V_n is denoted by $\bar{F}$.

(31) This clause may obviously be replaced by "V_n is conformal to $\bar{V}_n$ by means of a transformation of class C^m."

neighborhoods of $\bar{V}_n$ according to (2.3) considered as an admissible coordinate transformation, points in $\bar{V}_n$ and V_n with the same coordinates correspond and the conformal transformation becomes (2.5). Throughout this paper, unless a contrary assumption is explicitly made, we shall always assume that coordinate systems have been chosen so that (2.5) holds. In these coordinate systems,

$$\bar{g}_{ij} = e^{2\sigma} g_{ij}, \qquad \bar{g}^{ij} = e^{-2\sigma} g^{ij} \tag{2.6}$$

where g^{ij} and $\bar{g}^{ij}$ are the contravariant components of the metric tensors. Conversely, if (2.5) is a point transformation of the points of V_n and $\bar{V}_n$ and (2.6) holds at corresponding points where the mapping function $\sigma(x^i)$ is a real function of class C^{m-1}, it follows that $\bar{V}_n$ is conformal to V_n.

The problem of the conformal equivalence of Riemann spaces leads quite naturally to the study of the conformal Riemann space $\boldsymbol{V}_n$. The conformal Riemann space $\boldsymbol{V}_n$ of class C^m is a space whose coordinate manifold is of class C^m and whose fundamental geometric object, defined over the manifold, is the set of all second order, symmetric, positive definite tensors of class C^{m-1},

$$\{e^{2\sigma} g_{ij}\}, \tag{2.7}$$

any two of which are equal except for a positive multiplicative scalar factor of class C^{m-1}. The conformal tensor $g_{ij}/g^{1/n}$ constructed from any tensor g_{ij} belonging to (2.7) is independent of the particular tensor which is chosen. For this reason, T. Y. Thomas([32]) has defined the conformal Riemann space $\boldsymbol{V}_n$ by using this tensor instead of the set (2.7) as the fundamental geometric object of $\boldsymbol{V}_n$.

It is natural to associate the set of all conformally equivalent Riemann spaces

$$\{\bar{V}_n\} \tag{2.8}$$

of class C^m whose metric tensors (in some allowable coordinate system) belong to (2.7) with the conformal Riemann space $\boldsymbol{V}_n$. Indeed, as is easy to see, the geometric properties of $\boldsymbol{V}_n$ (which are independent of the factor $e^{2\sigma}$) are conformal properties of the set of Riemann spaces (2.8). Throughout this paper, whenever we refer to the conformally equivalent Riemann spaces V_n, $\bar{V}_n$, it will be understood that these spaces are *any* two spaces of the set of conformally equivalent Riemann spaces (2.8).

The above discussion shows that, formally, our conformal theory of curves is the theory, under the identity transformation, of a curve and an enveloping coordinate manifold on which is defined a second order, symmetric, positive definite tensor up to a positive scalar multiplicative factor. It is shown in §15, that our results may be used to develop a conformal theory of curves which is based directly on the conformal tensor $g_{ij}/g^{1/n}$.

([32]) T. Y. Thomas, loc. cit., p. 119.

Let $T^{i_1\cdots i_w}_{j_1\cdots j_v}$ be components of a tensor at a point P of V_n whose values depend upon geometric objects of V_n and of its subspaces [33]. Let $\bar{V}_n$ be any Riemann space conformal to V_n and let $\bar{T}^{i_1\cdots i_w}_{j_1\cdots j_v}$ be the components of the tensor at $\bar{P}$ whose values depend in the same way upon the corresponding geometric objects of $\bar{V}_n$ and its corresponding subspaces. Then if (2.6) holds and

$$\bar{T}^{i_1\cdots i_w}_{j_1\cdots j_v} = (e^{\sigma})^u T^{i_1\cdots i_w}_{j_1\cdots j_v} \tag{2.9}$$

we call $T^{i_1\cdots i_w}_{j_1\cdots j_v}$ a ***relative conformal tensor of weight*** u. The law of transformation of $T^{i_1\cdots i_w}_{j_1\cdots j_v}$ between any two $\bar{V}_n$ as well as between any two coordinate systems is consistent. If $u=0$, the tensor has the same components in V_n and $\bar{V}_n$. In this case, we call $T^{i_1\cdots i_w}_{j_1\cdots j_v}$ a *conformal tensor*. If $u=v-w$, we say that $T^{i_1\cdots i_w}_{j_1\cdots j_v}$ is a *conformetric tensor*. As will be seen below, these latter tensors have both metric and conformal properties.

Under the assumption that a relative conformal scalar Q exists in V_n, one may construct a conformetric tensor or a conformal tensor corresponding to every relative conformal tensor. More generally, under the same assumption, if $T^{i_1\cdots i_w}_{j_1\cdots j_v}$ obeys (2.9), one may construct a corresponding relative conformal tensor which satisfies an equation like (2.9) with u replaced by an arbitrary u'. For suppose that the transformation law[34] of Q is $\bar{Q}=e^{\sigma}Q$. Then $Q^{u'-u}\cdot T^{i_1\cdots i_w}_{j_1\cdots j_v}$ is a relative conformal tensor which satisfies (2.9) with u replaced by u'. We note that every relative conformal tensor (including conformetric tensors) is the product of a conformal tensor by a relative conformal scalar.

As a consequence of our definitions it follows that if the components of a relative conformal tensor are zero in V_n, they are zero in any $\bar{V}_n$. This fact permits us to write conformal tensor equations which retain their meaning under conformal transformations. The sum, difference, inner and outer product of conformetric tensors (conformal tensors) is also a conformetric tensor (conformal tensor).

If λ^i is a conformetric contravariant vector, the condition (2.9) becomes $\bar{\lambda}^i=e^{-\sigma}\lambda^i$. It follows that the direction of λ^i in V_n coincides with the direction of $\bar{\lambda}^i$ in $\bar{V}_n$. Since $\bar{g}_{ij}\bar{\lambda}^i\bar{\lambda}^j=g_{ij}\lambda^i\lambda^j$, the length of λ^i remains unchanged under any conformal mapping. Thus λ^i has a conformally invariant direction and (metric) length. Conversely, any vector for which this is true must be a conformetric vector. If the length of a conformetric vector is unity, then the vector is called a ***unit conformetric vector***. Any conformetric scalar is a conformal scalar or invariant. It is easy to show that any conformetric tensor

(33) Examples of such geometric objects which will be used in this paper are: the metric tensor g_{ij}, the Christoffel symbols of the second kind, the unit tangent and principal normal of a curve.

(34) There is no loss of generality in this assumption, for if $\bar{Q}=(e^{\sigma})^t\cdot Q$, the relative conformal scalar $|Q|^{1/t}$ has the desired transformation law.

which is not a scalar (and only these tensors) may be represented in the usual way[35] by means of adjoint n-beins of conformetric vectors. As follows from (2.6), g_{ij} and g^{ij} are conformetric tensors. Hence the indices of any conformetric tensor (but not of a conformal tensor) may be raised or lowered using g_{ij}, g^{ij} in the usual way and the result will be a conformetric tensor.

In this paper, we consider the conformal geometry of a curve in V_n. The curve introduces additional structure into V_n, by means of which a relative conformal scalar is found at points of the curve. In view of the existence of this relative conformal scalar, one may find a conformal vector corresponding to any conformetric vector, and conversely. Thus it is chiefly a matter of convenience whether we use conformetric vectors or conformal vectors. Our work is based upon unit conformetric vectors. The "conformal derivative" of such vectors is somewhat simpler than the "conformal derivative" of conformal vectors. However, we note that the analogous conformal theory of any subspace whose dimensionality exceeds one is developed by the use of conformal tensors.

3. **The conformal derivative.** We suppose that the class, defined in §2, of any two Riemann spaces V_n and $\overline{V}_n$ belonging to (2.8), is at least 2, that is, $m \geqq 2$. Then it follows from (2.6) that [36]

$$\overline{\left\{ {i \atop jk} \right\}} = \left\{ {i \atop jk} \right\} + \delta^i_j \sigma_{,k} + \delta^i_k \sigma_{,j} - g_{jk} g^{ih} \sigma_{,h} \tag{3.1}$$

where

$$\left\{ {i \atop jk} \right\}, \qquad \overline{\left\{ {i \atop jk} \right\}}$$

are the Christoffel symbols of the second kind for V_n and $\overline{V}_n$ respectively. Let

$$x^i = x^i(z)$$

represent a real curve C in V_n. Let $dx^i/dz \neq 0$ at each point of this z-interval for at least one value of i. We also suppose that the functions $x^i(z)$ are of class C^p where p is a fixed integer subject to the inequalities

$$m \geqq p \geqq 2. \tag{3.2}$$

Then it is easy to show that $s = s(z)$ is an allowable change of parameter where s is an arc length parameter determined up to an additive constant and a choice of sign. Hence the equation of C may be written as[37]

---

(35) A. Duschek and W. Mayer, *Lehrbuch der Differentialgeometrie*, vol. 2 (1930), pp. 14–15.

(36) The comma denotes covariant differentiation with respect to the x's and the form (2.1) and the δ^i_j are the Kronecker deltas.

(37) Note that $x^i(z)$ and $x^i(s)$ are different functions of their respective variables. This remark also applies to the functions $\bar{x}^i(z)$ and $\bar{x}^i(\bar{s})$ which are defined below.

$$x^i = x^i(s), \qquad a_1 < s < b_1,$$

where $x^i(s)$ are of class C^p. Then the unit tangent ν^i and the principal normal μ^i are given by

$$\nu^i = \frac{dx^i}{ds}, \tag{3.3}$$

$$\mu^i = \frac{d^2x^i}{ds^2} + \left\{ {i \atop jk} \right\} \frac{dx^j}{ds} \frac{dx^k}{ds}. \tag{3.4}$$

The curve $\overline{C}$ in $\overline{V}_n$ which corresponds to C under the conformal transformation (2.5) is

$$\bar{x}^i = \bar{x}^i(z)$$

where $\bar{x}^i(z)$ and $x^i(z)$ are the same functions of z and corresponding points have the same value of z. The conditions (3.2) also apply to $\overline{C}$. As shown above in the case of C, the curve $\overline{C}$ may also be referred to an arc length parameter $\bar{s}$ and written as

$$\bar{x}^i = \bar{x}^i(\bar{s}), \qquad \bar{a}_1 < \bar{s} < \bar{b}_1.$$

Naturally the points for which $s = \bar{s}$ do not correspond since metric arc length is not a conformal parameter. The unit tangent $\bar{\nu}^i$ and the principal normal $\bar{\mu}^i$ of $\overline{C}$ are given by equations similar to (3.3) and (3.4). From these equations and (2.4), (2.6), (3.1), (3.3) and (3.4), we find that

$$\bar{\nu}^i = e^{-\sigma}\nu^i, \tag{3.5}$$

$$\bar{\mu}^i = e^{-2\sigma}\left[\mu^i - \sigma_{,h}(g^{ih} - \nu^i\nu^h)\right]. \tag{3.6}$$

The tensor $g^{ih} - \nu^i\nu^h$ is the projection tensor[38] of the vector space orthogonal to ν^i. If we write μ_i and $\bar{\mu}_i$ for the covariant components of the principal normals, $\mu_i = g_{ij}\mu^j$, $\bar{\mu}_i = \bar{g}_{ij}\bar{\mu}^j$ and it follows from (2.6) and (3.6) that

$$\bar{\mu}_i = \mu_i - \sigma_{,i} + \sigma_{,k}\nu^k\nu_i \tag{3.7}$$

where ν_i is the covariant tangent defined by $\nu_i = g_{ij}\nu^j$.

Let $\lambda^i(t)$ be the components of a conformetric contravariant vector of class C^1 defined along C where t is any (not necessarily allowable) conformal parameter along C related to z (or s) by a parameter transformation of class C^1. Then dx^i/dt exists and is continuous. Since λ^i is a conformetric vector,

$$\bar{\lambda}^i = e^{-\sigma}\lambda^i. \tag{3.8}$$

We write the absolute derivative with respect to t and the form (2.1) of this vector as $D\lambda^i/Dt$ so that

(38) Duschek-Mayer, loc. cit., pp. 44–45.

$$\frac{D\lambda^i}{Dt} = \frac{d\lambda^i}{dt} + \left\{ {i \atop hk} \right\} \lambda^h \frac{dx^k}{dt}. \tag{3.9}$$

If we write the analogous equation for $\overline{D}\bar{\lambda}^i/\overline{D}t$ and simplify by the use of (3.1), (3.8) and (3.9), we obtain

$$\frac{\overline{D}\bar{\lambda}^i}{\overline{D}t} = e^{-\sigma}\left[\frac{D\lambda^i}{Dt} + \sigma_{,h}\lambda^h\frac{dx^i}{dt} - g_{hk}\lambda^h\frac{dx^k}{dt}g^{ij}\sigma_{,j}\right]. \tag{3.10}$$

We substitute the values for $\sigma_{,h}$ and $\sigma_{,j}g^{ij}$ which are obtained from (3.6) and (3.7) in (3.10) and simplify the resulting equation by the use of (2.4), (2.6), (3.3) and (3.8). This gives

$$\frac{\overline{D}\bar{\lambda}^i}{\overline{D}t} + \bar{\mu}_h\bar{\lambda}^h\frac{dx^i}{dt} - \bar{g}_{hk}\bar{\lambda}^h\frac{dx^k}{dt}\bar{\mu}^i = e^{-\sigma}\left[\frac{D\lambda^i}{Dt} + \mu_h\lambda^h\frac{dx^i}{dt} - g_{hk}\lambda^h\frac{dx^k}{dt}\mu^i\right].$$

It follows that $\mathfrak{d}\lambda^i/\mathfrak{d}t_C$ given by

$$\frac{\mathfrak{d}\lambda^i}{\mathfrak{d}t_C} = \frac{D\lambda^i}{Dt} + \mu_h\lambda^h\frac{dx^i}{dt} - g_{hk}\lambda^h\frac{dx^k}{dt}\mu^i \tag{3.11}$$

is a conformetric vector, that is,

$$\frac{\bar{\mathfrak{d}}\bar{\lambda}^i}{\bar{\mathfrak{d}}t_{\bar{C}}} = e^{-\sigma}\frac{\mathfrak{d}\lambda^i}{\mathfrak{d}t_C}.$$

The subscript C is used in the symbol $\mathfrak{d}/\mathfrak{d}t_C$ to indicate that the definition of this symbol depends upon the curve C as well as the process of differentiation. Since $\mathfrak{d}/\mathfrak{d}t_C$ will always be evaluated with respect to the same curve C in V_n, we shall usually write $\mathfrak{d}/\mathfrak{d}t$ for $\mathfrak{d}/\mathfrak{d}t_C$ without danger of ambiguity.

In order to arrive at a meaning for the operator $\mathfrak{d}/\mathfrak{d}t$ when applied to any tensor, we assume that $\mathfrak{d}/\mathfrak{d}t$ satisfies the following requirements[19]:

$$\frac{\mathfrak{d}\phi}{\mathfrak{d}t} = \frac{D\phi}{Dt} \tag{α}$$

if ϕ is any scalar;

$$\frac{\mathfrak{d}\lambda^i}{\mathfrak{d}t} = \frac{D\lambda^i}{Dt} + \Omega^i_j\lambda^j\frac{ds}{dt} \tag{β}$$

where λ^i is any contravariant vector (not necessarily a conformetric vector) and $\Omega^i_j = \nu^i\mu_j - \mu^i\nu_j$;

---

(19) This discussion is analogous to a similar one for ordinary covariant differentiation by Mayer. Cf. Duschek-Mayer, loc. cit., vol. 2, pp. 31–33.

$$(\gamma) \qquad \frac{\mathfrak{d}}{\mathfrak{d}t}(S\cdot T) = \frac{\mathfrak{d}S}{\mathfrak{d}t}\cdot T + S\cdot\frac{\mathfrak{d}T}{\mathfrak{d}t}$$

where S and T are any tensors

$$(\delta) \qquad \sum_{i=j}\left(\frac{\mathfrak{d}}{\mathfrak{d}t}T^i_j\right) = \frac{\mathfrak{d}}{\mathfrak{d}t}\left(\sum_{i=j}T^i_j\right)$$

where i and j are any two indices of the tensor T, one contravariant and the other covariant. (That is, the contraction operation for tensors $\sum_{i=j}$ and the $\mathfrak{d}/\mathfrak{d}t$ operation are commutative.)

On the basis of these properties, we shall find a unique expression for $\mathfrak{d}T/\mathfrak{d}t$. We first consider a covariant vector ξ_i. Then, according to (α),

$$\frac{\mathfrak{d}}{\mathfrak{d}t}(\lambda^i\xi_i) = \frac{D}{Dt}(\lambda^i\xi_i).$$

Because of (γ) and (δ), this may be written as

$$\frac{\mathfrak{d}\lambda^i}{\mathfrak{d}t}\cdot\xi_i + \lambda^i\frac{\mathfrak{d}\xi_i}{\mathfrak{d}t} = \frac{D\lambda^i}{Dt}\xi_i + \lambda^i\frac{D\xi_i}{Dt}\cdot$$

It follows that $\lambda^i(\mathfrak{d}\xi_i/\mathfrak{d}t)$ is an invariant for all λ^i and hence $\mathfrak{d}\xi_i/\mathfrak{d}t$ is a covariant vector (if it exists). In this last equation, we substitute the value for $\mathfrak{d}\lambda^i/\mathfrak{d}t$ given by (β) and simplify. This gives

$$\lambda^i\left(\frac{\mathfrak{d}\xi_i}{\mathfrak{d}t} - \frac{D\xi_i}{Dt} + \Omega^j_i\xi_j\frac{ds}{dt}\right) = 0.$$

Since λ^i is an arbitrary vector, it follows that

$$\frac{\mathfrak{d}\xi_i}{\mathfrak{d}t} = \frac{D\xi_i}{Dt} - \Omega^j_i\xi_j\frac{ds}{dt}$$

or

$$(3.12) \qquad \frac{\mathfrak{d}\xi_i}{\mathfrak{d}t_C} = \frac{D\xi_i}{Dt} - \xi_h\frac{dx^h}{dt}\mu_i + \xi_h\mu^h g_{ij}\frac{dx^j}{dt}\cdot$$

It is clear that $\mathfrak{d}/\mathfrak{d}t$ as applied to covariant vectors ξ_i satisfies those conditions (α) to (δ) which have meaning in this case.

To extend this definition to a tensor of any kind $T^{i_1\cdots i_w}_{j_1\cdots j_v}$, we form the invariant

$$(3.13) \qquad T^{i_1\cdots i_w}_{j_1\cdots j_v}\,{}_{(1)}\lambda^{j_1}\cdots{}_{(v)}\lambda^{j_v}\,{}_{(1)}\xi_{i_1}\cdots{}_{(w)}\xi_{i_w}$$

where the λ^j and ξ_i are arbitrary vectors. If we apply the $\mathfrak{d}/\mathfrak{d}t$ operator to

(3.13) and proceed as in the derivation of (3.12), using (α) to (δ) and (3.12), we find that

$$\begin{aligned}(3.14)\qquad \frac{\mathfrak{d}}{\mathfrak{d}t_C} T^{i_1\cdots i_w}_{j_1\cdots j_v} = \frac{D}{Dt} T^{i_1\cdots i_w}_{j_1\cdots j_v} + \sum_{\alpha=1}^{w} T^{i_1\cdots i_{\alpha-1}hi_{\alpha+1}\cdots i_w}_{j_1\cdots j_v} \Omega^{i_\alpha}_h \frac{ds}{dt} \\ - \sum_{\beta=1}^{v} T^{i_1\cdots i_w}_{j_1\cdots j_{\beta-1}hj_{\beta+1}\cdots j_v} \Omega^h_{j_\beta} \frac{ds}{dt}.\end{aligned}$$

It follows that $\mathfrak{d}T^{i_1\cdots i_w}_{j_1\cdots j_v}/\mathfrak{d}t$ is a tensor of the same kind as $T^{i_1\cdots i_w}_{j_1\cdots j_v}$ and that the definition of $\mathfrak{d}T/\mathfrak{d}t$ stated in (3.14) satisfies conditions (α) to (δ). Equations (3.11) and (3.12) are special cases of (3.14).

The definition of $\mathfrak{d}T/\mathfrak{d}t$ stated in (3.14) may be based upon symmetric coefficients of connection Γ^i_{jk} just as ordinary covariant differentiation is based upon the Christoffel symbols $\left\{ {i \atop jk} \right\}$. We define the Γ^i_{jk} by[40]

$$\Gamma^i_{jk} = \left\{ {i \atop jk} \right\} + \mu_j\delta^i_k + \mu_k\delta^i_j - g_{jk}\mu^i$$

and note that (3.14) is equivalent to

$$\begin{aligned}\frac{\mathfrak{d}}{\mathfrak{d}t} T^{i_1\cdots i_w}_{j_1\cdots j_v} = \frac{d}{dt} T^{i_1\cdots i_w}_{j_1\cdots j_v} + \sum_{\alpha=1}^{w} T^{i_1\cdots i_{\alpha-1}hi_{\alpha+1}\cdots i_w}_{j_1\cdots j_v} \Gamma^{i_\alpha}_{hk} \frac{dx^k}{dt} \\ - \sum_{\beta=1}^{v} T^{i_1\cdots i_w}_{j_1\cdots j_{\beta-1}hj_{\beta+1}\cdots j_v} \Gamma^h_{j_\beta k} \frac{dx^k}{dt}.\end{aligned}$$

We now find the law of transformation of $\mathfrak{d}T^{i_1\cdots i_w}_{j_1\cdots j_v}/\mathfrak{d}t$ when $T^{i_1\cdots i_w}_{j_1\cdots j_v}$ is a relative conformal tensor of weight u; that is, when $T^{i_1\cdots i_w}_{j_1\cdots j_v}$ obeys (2.9). As follows from the definition of Γ^i_{jk} and (2.6), (3.1), (3.6) and (3.7),

$$\overline{\Gamma}^i_{jk} = \Gamma^i_{jk} + \frac{d\sigma}{ds}(\nu_j\delta^i_k + \nu_k\delta^i_j - g_{jk}\nu^i).$$

By means of this relation and the definition for $\mathfrak{d}T^{i_1\cdots i_w}_{j_1\cdots j_v}/\mathfrak{d}t$ by means of the Γ^i_{jk}, we find upon applying the $\mathfrak{d}/\mathfrak{d}t$ operator to (2.9) that

$$\frac{\overline{\mathfrak{d}}}{\mathfrak{d}t} \overline{T}^{i_1\cdots i_w}_{j_1\cdots j_v} = e^{\sigma u}\left\{ \frac{\mathfrak{d}}{\mathfrak{d}t} T^{i_1\cdots i_w}_{j_1\cdots j} + (u + w - v)\frac{d\sigma}{dt} T^{i_1\cdots i_w}_{j_1\cdots j_v} \right\}.$$

---

[40] Since the Γ^i_{jk} differ from the Christoffel symbols

$$\left\{ {i \atop jk} \right\}$$

by a tensor, it is immediate that they must transform like coefficients of connection under coordinate transformations. For example, cf. L. P. Eisenhart, *Non-Riemannian Geometry*, American Mathematical Society Colloquium Publications, vol. 8, 1927, p. 48.

If Q is a relative conformal scalar of weight -1 so that $\overline{Q}=e^{-\sigma}Q$, the last equations may be written in invariant form by noting that they are equivalent to the equations

$$S^{i_1\cdots i_u}_{j_1\cdots j_v} = \frac{\mathfrak{d}}{\mathfrak{d}t} T^{i_1\cdots i_u}_{j_1\cdots j_v} + (u + w - v)\left(\frac{d \log Q}{dt}\right) T^{i_1\cdots i_u}_{j_1\cdots j_v},$$

$$\overline{S}^{i_1\cdots i_u}_{j_1\cdots j_v} = e^{\sigma u} S^{i_1\cdots i_u}_{j_1\cdots j_v}.$$

Hence $S^{i_1\cdots i_u}_{j_1\cdots j_v}$ is a relative conformal tensor of weight u. In particular, if $T^{i_1\cdots i_u}_{j_1\cdots j_v}$ is a conformetric tensor so that $u=v-w$, it follows that $\mathfrak{d}T^{i_1\cdots i_u}_{j_1\cdots j_v}/\mathfrak{d}t$ is also a conformetric tensor. This fact exhibits the conformal property of the operator $\mathfrak{d}/\mathfrak{d}t$ and thus justifies the definition: *The tensor* $\mathfrak{d}T^{i_1\cdots i_u}_{j_1\cdots j_v}/\mathfrak{d}t$ *defined by* (3.14) *is called the conformal derivative* (*with respect to the curve* C) *of* $T^{i_1\cdots i_u}_{j_1\cdots j_v}$ *with respect to* t.

The conformal derivative at a point is thus dependent not only on the metric of V_n but also on the curve C (or rather, on the second order element of C). This dependence of differentiation on a curve as well as the space is analogous to the similar dependence of parallel displacement of vectors in a general Riemann space. In this respect the conformal derivative also resembles the derivative (with respect to a curve) which has been defined for any Finsler space by Synge([41]) and Taylor([42]). Indeed, it is very likely that the results of this paper may be generalized to apply to any Finsler space.

The geometry which is based upon the conformal derivative will appear in a separate paper. We note here a number of fundamental properties of the conformal derivative which are elementary consequences of the preceding remarks and the definition:

(A) *The conformal derivative with respect to a conformal parameter of a conformetric tensor is a conformetic tensor.*

(B) *The conformal derivative of any tensor is a tensor.*

(C) *Conformal differentiation of the sum, difference, inner and outer product of tensors obeys the same rules as ordinary differentiation.*

(D) *The conformal derivative of* g_{ij}, g^{ij}, δ^i_j *is zero; that is,*

$$\frac{\mathfrak{d}g_{ij}}{\mathfrak{d}t} = \frac{\mathfrak{d}g^{ij}}{\mathfrak{d}t} = \frac{\mathfrak{d}\delta^i_j}{\mathfrak{d}t} = 0. \tag{3.15}$$

(E) *The conformal derivative* (*with respect to a curve* C) *of the unit tangent vector of* C *is zero, that is,*

$$\frac{\mathfrak{d}\nu^i}{\mathfrak{d}t} = 0. \tag{3.16}$$

([41]) J. L. Synge, *A generalization of the Riemannian line element*, these Transactions, vol. 27 (1925), p. 64.

([42]) J. H. Taylor, *A generalization of Levi-Civita's parallelism and the Frenet formulas*, these Transactions, vol. 27 (1925), pp. 255–257.

The conformal derivative (3.14) is with respect to t which is a conformal parameter, that is, corresponding points of C and $\bar{C}$ have the same value of t. The equations of C and $\bar{C}$ are frequently given in terms of metric arc length parameters s and $\bar{s}$. In this case, the conformal derivative of $T^{i_1\cdots i_p}_{j_1\cdots j_q}$ with respect to s is given by (3.14) with t replaced by s. It follows that

$$\frac{\mathfrak{d}}{\mathfrak{d}s} T^{i_1\cdots i_p}_{j_1\cdots j_q} = \frac{dt}{ds}\cdot\frac{\mathfrak{d}}{\mathfrak{d}t} T^{i_1\cdots i_p}_{j_1\cdots j_q}. \tag{3.17}$$

Since

$$\frac{dt}{d\bar{s}} = e^{-\sigma}\frac{dt}{ds}, \tag{3.18}$$

dt/ds is a relative conformal scalar and it follows from (3.17) that if $T^{i_1\cdots i_p}_{j_1\cdots j_q}$ is a conformetric tensor then $\mathfrak{d}T^{i_1\cdots i_p}_{j_1\cdots j_q}/\mathfrak{d}s$ is the product of a relative conformal scalar by a conformetric tensor. The conformal derivative with respect to s has the properties (B), (C), (D), and (E) mentioned above. We note that the existence of a conformal parameter t implies the existence of a relative conformal scalar. For according to (3.18), dt/ds is a relative conformal scalar. Conversely, let Q be a relative conformal scalar which we may assume transforms so that $\bar{Q}=e^{-\sigma}Q$. Then the solution t of the differential equation $dt/ds=Q(s)$ is a conformal parameter.

A simple geometric interpretation of $\mathfrak{d}\lambda^i/\mathfrak{d}t$ is possible in terms of the ideas of projection and ordinary (metric) differentiation. We denote by $N_\phi\lambda^i$ the projection of λ^i in the vector space normal to an arbitrary vector ϕ^i. Then if ψ^i, ψ_i are the unit contravariant and covariant vectors which span the vector space determined by ϕ^i

$$N_\phi\lambda^i = \lambda^i - \lambda^h\psi_h\cdot\psi^i.$$

Let

$$\omega^i = N_\nu\lambda^i \tag{3.19}$$

where ν^i is the unit tangent to the curve C. It follows that

$$\lambda^i = \omega^i + \alpha\nu^i$$

where

$$\omega^i\nu_i = 0 \tag{3.20}$$

and α is a scalar. Then

$$\frac{\mathfrak{d}\lambda^i}{\mathfrak{d}t} = \frac{\mathfrak{d}\omega^i}{\mathfrak{d}t} + \frac{d\alpha}{dt}\nu^i + \alpha\frac{\mathfrak{d}\nu^i}{\mathfrak{d}t}$$

according to (C). According to (3.11), (3.16), (3.20) and the last equation,

$$\frac{\mathfrak{d}\lambda^i}{\mathfrak{d}t} = \frac{D\omega^i}{Dt} + \mu_h\omega^h\frac{dx^i}{dt} + \frac{d\alpha}{dt}\nu^i.$$

By absolute differentiation of (3.20) with respect to t and the form (2.1) and use of (3.3) and (3.4), we obtain

$$\frac{D\omega^h}{Dt}\nu_h + \omega^h\mu_h\frac{ds}{dt} = 0.$$

Now

$$N_\nu\frac{D\omega^i}{Dt} = \frac{D\omega^i}{Dt} - \frac{D\omega^h}{Dt}\nu_h\cdot\nu^i.$$

It follows from (3.19) and the last three equations that

$$\frac{\mathfrak{d}\lambda^i}{\mathfrak{d}t} = N_\nu\frac{D}{Dt}N_\nu\lambda^i + \frac{d\alpha}{dt}\nu^i \tag{3.21}$$

which is the desired interpretation. In particular, if λ^i is orthogonal to the curve C then $\alpha=0$ and $N_\nu\lambda^i=\lambda^i$ so that

$$\frac{\mathfrak{d}\lambda^i}{\mathfrak{d}t} = N_\nu\frac{D\lambda^i}{Dt}. \tag{3.22}$$

It follows that if λ^i is orthogonal to C then $\mathfrak{d}\lambda^i/\mathfrak{d}t$ is also orthogonal to C.

It is also possible to define a derivative which has the property that when applied to a conformal vector it yields a conformal vector. If we suppose that λ^i is a conformal vector of class C^1 then

$$\bar{\lambda}^i = \lambda^i. \tag{3.23}$$

We proceed with this equation precisely as with (3.8) and find that

$$\frac{\overline{\mathfrak{d}\lambda^i}}{\overline{\mathfrak{d}t}} = \frac{\mathfrak{d}\lambda^i}{\mathfrak{d}t} + \alpha\lambda^i$$

where $\mathfrak{d}\lambda^i/\mathfrak{d}t$ is defined by (3.11) and α is a scalar. It follows from this equation that

$$\frac{\mathcal{D}\lambda^i}{\mathcal{D}t_C} = N_\lambda\frac{\mathfrak{d}\lambda^i}{\mathfrak{d}t_C} \tag{3.24}$$

is a conformal vector. In view of the more complicated structure of $\mathcal{D}\lambda^i/\mathcal{D}t_C$, our work shall be in terms of conformetric vectors and the conformal derivative.

Another type of differentiation for which the derivative of a conformal vector λ^i of class C^1 is a conformal vector may be defined if a nonzero rela-

tive conformal scalar Q of class C^1 exists along the curve. We may assume that the law of transformation of Q is

$$\bar{Q} = e^{-\sigma} Q. \tag{3.25}$$

Then in virtue of (3.23) and (3.25), $Q\lambda^i$ is a conformetric vector and, in accordance with (A), $\mathfrak{d}(Q\lambda^i)/\mathfrak{d}t$ is also a conformetric vector. This fact and (3.25) show that $\delta\lambda^i/\delta t_{(C,Q)}$ defined by

$$\frac{\delta\lambda^i}{\delta t_{(C,Q)}} = Q^{-1}\frac{\mathfrak{d}Q\lambda^i}{\mathfrak{d}t_C} = \frac{d\log Q}{dt}\cdot\lambda^i + \frac{\mathfrak{d}\lambda^i}{\mathfrak{d}t_C} \tag{3.26}$$

is a conformal vector. The type of differentiation defined for contravariant vectors by (3.26) may be extended to conformal tensors $T^{i_1\cdots i_v}_{j_1\cdots j_w}$. One readily finds by reasoning similar to that used in the derivation of (3.26) that $\delta T^{i_1\cdots i_v}_{j_1\cdots j_w}/\delta t_{(C,Q)}$ defined by

$$\begin{aligned}\frac{\delta}{\delta t_{(C,Q)}} T^{i_1\cdots i_v}_{j_1\cdots j_w} &= Q^{v-w}\frac{\mathfrak{d}}{\mathfrak{d}t_C}\left(Q^{w-v}\cdot T^{i_1\cdots i_v}_{j_1\cdots j_w}\right)\\ &= (w-v)\frac{d\log Q}{dt}\cdot T^{i_1\cdots i_v}_{j_1\cdots j_w} + \frac{\mathfrak{d}}{\mathfrak{d}t_C} T^{i_1\cdots i_v}_{j_1\cdots j_w}\end{aligned} \tag{3.27}$$

is a conformal tensor. As a consequence of (3.24) and (3.26), we note that

$$\frac{\mathcal{D}\lambda^i}{\mathcal{D}t_C} = N_\lambda\frac{\delta\lambda^i}{\delta t_{(C,Q)}}\cdot \tag{3.28}$$

While the results of this paper are based upon the $\mathfrak{d}/\mathfrak{d}t$ process, they may be derived equally well using the $\delta/\delta t$ type of differentiation defined by (3.27).

If $T^{i_1\cdots i_v}_{j_1\cdots j_w}$ is any tensor (not necessarily a conformal tensor), we define $\delta T^{i_1\cdots i_v}_{j_1\cdots j_w}/\delta t_{(C,Q)}$ by means of (3.27). Since $\delta T^{i_1\cdots i_v}_{j_1\cdots j_w}/\delta t_{(C,Q)}$ will always be evaluated with respect to the same curve C, we may write $\delta/\delta t_Q$ for $\delta/\delta t_{(C,Q)}$. We easily prove that the properties (B) and (C) hold for the $\delta/\delta t_Q$ type of conformal differentiation and that (A), (D), and (E) are replaced by the analogous statements:

(A′) *The $\delta/\delta t_Q$ derivative of a conformal tensor is a conformal tensor.*

(D′) *The $\delta/\delta t_Q$ derivative of $Q^2 g_{ij}$, $Q^{-2}g^{ij}$, δ^i_j is zero.*

(E′) *The $\delta/\delta t_{(C,Q)}$ derivative of the conformal vector $Q^{-1}\nu^i$ which corresponds to the unit tangent of C is zero.*

Just as $\mathfrak{d}/\mathfrak{d}t_C$ differentiation may be defined by means of the Γ^i_{jk}, so the definition of $\delta/\delta t_{(C,Q)}$ differentiation may be based upon symmetric coefficients of connection Γ'^i_{jk}. These Γ'^i_{jk} are defined along points of the curve by means of the equations[43]

---

[43] The footnote concerning the law of transformation of the Γ^i_{jk} also applies here so that the Γ'^i_{jk} transform like coefficients of connection under coordinate transformations.

$$\Gamma'^{i}_{jk} = \left\{ {i \atop jk} \right\} + \left(\mu_j + \nu_j \frac{d \log Q}{ds}\right)\delta^i_k + \left(\mu_k + \nu_k \frac{d \log Q}{ds}\right)\delta^i_j - g_{jk}\left(\mu^i + \nu^i \frac{d \log Q}{ds}\right).$$

Then a simple calculation shows that (3.27) is equivalent to

$$\frac{\delta}{\delta t_{(C,Q)}} T^{i_1 \cdots i_w}_{j_1 \cdots j_v} = \frac{dT^{i_1 \cdots i_w}_{j_1 \cdots j_v}}{dt} + \sum_{\alpha=1}^{w} T^{i_1 \cdots i_{\alpha-1} h i_{\alpha+1} \cdots i_w}_{j_1 \cdots j_v} \Gamma'^{i_\alpha}_{hk} \frac{dx^k}{dt} - \sum_{\beta=1}^{v} T^{i_1 \cdots i_w}_{j_1 \cdots j_{\beta-1} h j_{\beta+1} \cdots j_v} \Gamma'^{h}_{j_\beta k} \frac{dx^k}{dt}.$$

The law of transformation of the Γ'^{i}_{jk} under a conformal transformation of V_n is found by using the definition of Γ'^{i}_{jk} and equations (2.4), (2.6), (3.1), (3.5), (3.6), (3.7) and (3.25). It is

$$\bar{\Gamma}'^{i}_{jk} = \Gamma'^{i}_{jk}.$$

The conformal invariance of the Γ'^{i}_{jk} provides another proof of (A'). In the same way, it may be shown that if $T^{i_1 \cdots i_w}_{j_1 \cdots j_v}$ is a relative conformal tensor of weight u then

$$\frac{\delta}{\delta t_Q} T^{i_1 \cdots i_w}_{j_1 \cdots j_v} + u \frac{d \log Q}{dt} \cdot T^{i_1 \cdots i_w}_{j_1 \cdots j_v}$$

is also a relative conformal tensor of weight u. We remark that the conformal theory of a general subspace is based upon conformal tensors and a type of "conformal differentiation" analogous to the $\delta/\delta t_{(C,Q)}$ differentiation.

4. **The conformal Frenet equations.** We suppose that the inequalities (3.2) hold throughout this section unless it is stated otherwise. Let ${}_{(1)}\theta^i(t)$ where t is the conformal parameter defined in the previous section be an arbitrary unit conformetric vector of class C^r $(m-1 \geqq r \geqq 1)$ defined along C. Then the corresponding vector defined along $\bar{C}$ in $\bar{V}_n$ is given by ${}_{(1)}\bar{\theta}^i = e^{-\sigma}\, {}_{(1)}\theta^i$ and must also be a vector of class C^r. We shall derive conformal analogues of the ordinary Frenet equations for the vector ${}_{(1)}\theta^i$ subject to the assumption that the "normals" in these equations exist and are of class C^1. For the satisfaction of this last assumption it is sufficient but not necessary that[44]

(4.1) $$m \geqq p \geqq n+1, \qquad r \geqq n.$$

By conformal differentiation of ${}_{(1)}\theta^i$ with respect to t, we obtain $\mathfrak{d}\,{}_{(1)}\theta^i/\mathfrak{d}t$. If $\mathfrak{d}\,{}_{(1)}\theta^i/\mathfrak{d}t$ is not identically zero, at points where at least one component does

[44] We may weaken these assumptions by replacing n by the integer τ which is defined below.

not vanish we may write

$$\frac{\mathfrak{d}\ {}_{(1)}\theta^i}{\mathfrak{d}t} = H_1 \cdot {}_{(2)}\theta^i \tag{4.2}$$

where $H_1 \neq 0$ and ${}_{(2)}\theta^i$ is a unit vector, that is,

$$g_{ij}\ {}_{(2)}\theta^i\ {}_{(2)}\theta^j = 1. \tag{4.3}$$

Both H_1 and ${}_{(2)}\theta^i$ are determined except for an initial choice of sign.

By conformal differentiation with respect to t of

$$g_{ij}\ {}_{(1)}\theta^i\ {}_{(1)}\theta^j = 1 \tag{4.4}$$

and use of (3.15) and (4.2), we obtain

$$g_{ij}\ {}_{(1)}\theta^i\ {}_{(2)}\theta^j = 0. \tag{4.5}$$

We suppose that the class of ${}_{(2)}\theta^i$ is C^1. Then, since ${}_{(2)}\theta^i$ is a unit vector, $\mathfrak{d}\ {}_{(2)}\theta^i/\mathfrak{d}t$ is normal to ${}_{(2)}\theta^i$. If $\mathfrak{d}\ {}_{(2)}\theta^i/\mathfrak{d}t$ is not contained in the linear vector space determined by ${}_{(1)}\theta^i$, it may be written as

$$\frac{\mathfrak{d}\ {}_{(2)}\theta^i}{\mathfrak{d}t} = A_1\ {}_{(1)}\theta^i + H_2\ {}_{(3)}\theta^i \tag{4.6}$$

where ${}_{(3)}\theta^i$ is a unit vector normal to ${}_{(1)}\theta^i$ and ${}_{(2)}\theta^i$ and $H_2 \neq 0$. From (4.4) and (4.6) we find that $A_1 = g_{ij}(\mathfrak{d}\ {}_{(2)}\theta^i/\mathfrak{d}t)\ {}_{(1)}\theta^j$. Now, if we differentiate (4.5) conformally with respect to t and use (3.15), (4.2) and (4.3) we obtain

$$H_1 + g_{ij}\ {}_{(1)}\theta^i \frac{\mathfrak{d}\ {}_{(2)}\theta^j}{\mathfrak{d}t} = 0.$$

Hence (4.6) becomes

$$\frac{\mathfrak{d}\ {}_{(2)}\theta^i}{\mathfrak{d}t} = -H_1\ {}_{(1)}\theta^i + H_2\ {}_{(3)}\theta^i. \tag{4.7}$$

Equations (4.2) and (4.7) are the first two analogues of the Frenet equations([45]). We shall prove the general formula by means of the customary proof by mathematical induction. Suppose

$$\frac{\mathfrak{d}\ {}_{(\alpha)}\theta^i}{\mathfrak{d}t} = -H_{\alpha-1}\ {}_{(\alpha-1)}\theta^i + H_\alpha\ {}_{(\alpha+1)}\theta^i \tag{4.8}$$

with the convention([46]) $H_0 = 0$ where $\alpha = 1, 2, \cdots, l-1$ and where the vectors ${}_{(\beta)}\theta^i$ satisfy the relations

([45]) For the general metric Frenet equations cf. L. P. Eisenhart, *Riemannian Geometry*, 1926, pp. 103–107. Also cf. Duschek-Mayer, loc. cit., pp. 59–62.

([46]) If $l = n$, we also make the convention that $H_n = 0$.

$$g_{ij\ (\beta)}\theta^i\ _{(\gamma)}\theta^j = \delta^\beta_\gamma \tag{4.9}$$

where $\beta, \gamma = 1, 2, \cdots, l$. If the class of $_{(l)}\theta^i$ is C^1 we shall show that the equations (4.8) and (4.9) hold if the ranges of α, β, γ are each increased by one.

The conformal derivative $\mathfrak{d}\ _{(l)}\theta^i/\mathfrak{d}t$ has a representation of the form

$$\frac{\mathfrak{d}\ _{(l)}\theta^i}{\mathfrak{d}t} = A_1\ _{(1)}\theta^i + \cdots + A_l\ _{(l)}\theta^i + \omega^i \tag{4.10}$$

where $g_{ij\ (\beta)}\theta^i\omega^j = 0$ $(\beta = 1, 2, \cdots, l)$. As above, we find

$$A_\beta = g_{ij\ (\beta)}\theta^j \frac{\mathfrak{d}\ _{(l)}\theta^i}{\mathfrak{d}t}, \qquad \beta = 1, 2, \cdots, l.$$

As a result of (4.9), we find that

$$g_{ij}\frac{\mathfrak{d}\ _{(l)}\theta^i}{\mathfrak{d}t}\ _{(\beta)}\theta^j = -\ g_{ij}\frac{\mathfrak{d}\ _{(\beta)}\theta^i}{\mathfrak{d}t}\ _{(l)}\theta^j.$$

For $\beta = l$, this shows that $A_l = 0$. If we substitute for $\mathfrak{d}\ _{(\beta)}\theta^i/\mathfrak{d}t$ $(\beta = 1, 2, \cdots, l-2)$ from (4.8) and use (4.9), we find that $A_1 = A_2 = \cdots = A_{l-2} = 0$. In the same way, the work for $\beta = l-1$ shows that $A_{l-1} = -H_{l-1}$. Furthermore if ω^i is not identically a zero vector, then at points where at least one of its components does not vanish we may write $\omega^i = H_l \cdot {}_{(l+1)}\theta^i$ where $H_l \neq 0$ and $g_{ij\ (l+1)}\theta_i\ _{(l+1)}\theta^j = 1$. Then (4.10) becomes (4.8) with $\alpha = l$ and the lth formula holds. It is clear that (4.9) is also true for the greater range of β, γ.

This process of constructing successive θ^i, H may be continued until we arrive at a vector $_{(r)}\theta^i$ (whose class is assumed to be C^1) such that $\mathfrak{d}\ _{(r)}\theta^i/\mathfrak{d}t$ is contained in the linear vector space determined by $_{(1)}\theta^i, {}_{(2)}\theta^i, \cdots, {}_{(r)}\theta^i$. Then $H_r = 0$ by definition and (4.8) and (4.9) hold for $\alpha, \beta, \gamma = 1, 2, \cdots, r$. In this case, by definition, $H_{r+1} = \cdots = H_n = 0$. We shall sometimes write $_{(r+1)}\theta^i, \cdots, {}_{(n)}\theta^i$ for any vectors which obey (4.9). Of course, $r \leqq n$. We call (4.8) where the $_{(\alpha)}\theta^i$ obey (4.9) for $\alpha = 1, 2, \cdots, r$ the *conformal Frenet equations in V_n for the vector $_{(1)}\theta^i$ and the parameter t.*

If V_n is mapped conformally on $\overline{V}_n$, as already shown, $_{(1)}\bar{\theta}^i$ must exist and have at least one continuous derivative. Hence an equation corresponding to (4.2) in $\overline{V}_n$ exists and may be written as

$$\frac{\bar{\mathfrak{d}}\ _{(1)}\bar{\theta}^i}{\bar{\mathfrak{d}}t} = \overline{H}_1\ _{(2)}\bar{\theta}^i \tag{4.11}$$

where $_{(2)}\bar{\theta}^i$ is a unit vector. According to §3 (B), $\bar{\mathfrak{d}}\ _{(1)}\bar{\theta}^i/\bar{\mathfrak{d}}t = e^{-\sigma}(\mathfrak{d}\ _{(1)}\theta^i/\mathfrak{d}t)$. This last equation and (4.2) and (4.11) show that $_{(2)}\theta^i$ is a unit conformetric vector, that is, $_{(2)}\bar{\theta}^i = e^{-\sigma}\ _{(2)}\theta^i$ and that H_1 is a conformal scalar, that is, $\overline{H}_1 = H$. It follows that the classes of $_{(2)}\bar{\theta}^i$ and $_{(2)}\theta^i$ are equal.

Since the class of $_{(2)}\theta^i$ was assumed to be C^1, this must also be true of $_{(2)}\bar{\theta}^i$

so that the second conformal Frenet equation in $\overline{V}_n$ is valid. A comparison of this equation with (4.7) using the properties of the conformal derivative readily shows that ${}_{(2)}\theta^i$ is a unit conformetric vector and H_2 is a conformal scalar and that ${}_{(2)}\bar{\theta}^i$ and ${}_{(3)}\theta^i$ are of class C^1.

This method of reasoning may be continued and leads to the following conclusion: *If the conformal Frenet equations in V_n exist then the conformal Frenet equations in $\overline{V}_n$ also exist, that is,*

$$\frac{\bar{\delta}\ {}_{(\alpha)}\bar{\theta}^i}{\bar{\delta}t} = -\overline{H}_{\alpha-1}\ {}_{(\alpha-1)}\bar{\theta}^i + \overline{H}_{\alpha}\ {}_{(\alpha+1)}\bar{\theta}^i, \qquad \overline{H}_0 = \overline{H}_r = 0,$$

$$\bar{g}_{ij}\ {}_{(\beta)}\bar{\theta}^i\ {}_{(\gamma)}\bar{\theta}^j = \delta^\beta_\gamma$$

where $\alpha, \beta, \gamma = 1, 2, \cdots, r$. *In addition the* ${}_{(\alpha)}\theta^i$ *and* H_α *satisfy the equations*

$$(4.12) \qquad {}_{(\alpha)}\bar{\theta}^i = e^{-\sigma}\ {}_{(\alpha)}\theta^i, \qquad \overline{H}_\alpha = H_\alpha.$$

If t is replaced by another conformal parameter t' and we denote the quantities in the corresponding conformal Frenet equations([47]) by primes, then ${}_{(\alpha)}\theta'^i = {}_{(\alpha)}\theta^i$, $H'_\alpha = (dt/dt')\cdot H_\alpha$. Hence the ${}_{(\alpha)}\theta^i$ are independent of the parametrization and the H_α are multiplied by the same conformal scalar dt/dt'. In the same way if the metric arc length s is used instead of t, one finds that the ${}_{(\alpha)}\theta^i$ are unchanged and H_α are multiplied by the same relative conformal scalar dt/ds. We call $H_1, H_2, \cdots, H_{r-1}$ the *associate conformal curvatures of the vector* ${}_{(1)}\theta^i$ *and the parameter* t *of orders* $1, 2, \cdots, r-1$ and say that ${}_{(2)}\theta^i, {}_{(3)}\theta^i, \cdots, {}_{(r)}\theta^i$ are the *associate conformal directions of the vectors* ${}_{(1)}\theta^i$ *of orders* $1, 2, \cdots, r-1$. We also say that ${}_{(2)}\theta^i, {}_{(3)}\theta^i, \cdots, {}_{(r)}\theta^i$ are obtained from ${}_{(1)}\theta^i$ by the *Frenet process*.

When the vector ${}_{(1)}\theta^i$ is normal to C, $H_{n-1}\equiv 0$. For according to the discussion following (3.21) and the conformal Frenet equations, if $H_1H_2\cdots H_{n-2}\not\equiv 0$ then ${}_{(1)}\theta^i, {}_{(2)}\theta^i, \cdots, {}_{(n-1)}\theta^i$ are all normal to ν^i. Now, if $H_{n-1}\not\equiv 0$ then ${}_{(n)}\theta^i$ would also be orthogonal to ν^i. Since this is impossible in virtue of (4.9) it follows that $H_{n-1}\equiv 0$, ${}_{(n)}\theta^i = \nu^i$. If $H_1H_2\cdots H_{n-2}\equiv 0$ then $H_{n-1}\equiv 0$ by definition.

The familiar and most useful metric Frenet equations are those obtained when the first vector is the unit tangent ν^i and the parameter is a metric arc length parameter s. In this case, the equations are

$$(4.13) \qquad \frac{D\ {}_{(\alpha)}\nu^i}{Ds} = -k_{\alpha-1}\ {}_{(\alpha-1)}\nu^i + k_\alpha\ {}_{(\alpha+1)}\nu^i, \qquad \alpha = 1, 2, \cdots, n,$$

where we make the convention that $k_0 = k_n = 0$. The metric invariants $k_1, k_2, \cdots, k_{n-1}$ are the successive (metric) curvatures and ${}_{(1)}\nu^i\ (=\nu^i)$, ${}_{(2)}\nu^i, \cdots, {}_{(n)}\nu^i$ are the unit tangent and successive unit (metric) normals which obey the relations

([47]) These equations exist if the conformal Frenet equations for the parameter t exist.

$$g_{ij\,(\alpha)}\nu^{i}\,{}_{(\beta)}\nu^{j} = \delta_{\alpha}^{\beta}. \tag{4.14}$$

The conformal Frenet equations given by (4.8) hold for any first vector ${}_{(1)}\theta^i$ and any parameter t. In order to derive an exact analogue of (4.13), we must indicate a "natural" vector and a "natural" conformal parameter which will play roles analogous to the roles of ν^i and s respectively in the Frenet equations (4.13). We note that the obvious selection for ${}_{(1)}\theta^i$, namely ν^i, is an unfruitful one. For in view of (3.16), the first associate conformal curvature of ν^i vanishes identically.

In order to obtain a solution of this problem, for the remainder of this section we shall replace the assumption stated in (3.2) by the inequalities

$$m \geqq p \geqq 3. \tag{4.15}$$

Then the components of the vectors μ^i and $\bar{\mu}^i$ have continuous first derivatives. If we consider a conformetric vector whose components in $\bar{V}_n$ are $\bar{\mu}^i$ then, according to (3.6), the corresponding components in V_n are $e^{-\sigma}[\mu^i-\sigma_{,h}(g^{ih}-\nu^i\nu^h)]$. It follows from §3 (A) and (3.17) that

$$\frac{\bar{\delta}\bar{\mu}^i}{\bar{\delta}\bar{s}} = e^{-2\sigma}\left[\frac{\delta}{\delta s}\left\{e^{-\sigma}[\mu^i - \sigma_{,h}(g^{ih} - \nu^i\nu^h)]\right\}\right].$$

Since $\bar{\mu}^i$ and μ^i are normal to ν^i, (3.22) applies so that the last equation becomes

$$\bar{N}_\nu \frac{\bar{D}\bar{\mu}^i}{\bar{D}\bar{s}} = e^{-3\sigma}\left[N_\nu \frac{D\mu^i}{Ds} - \sigma_{hk}\nu^k(g^{ih} - \nu^i\nu^h)\right] \tag{4.16}$$

where

$$\sigma_{hk} = \sigma_{,hk} - \sigma_{,h}\sigma_{,k}. \tag{4.17}$$

Now according to the Frenet equations (4.13),

$$\mu^i = k_1\,{}_{(2)}\nu^i, \qquad \frac{D\,{}_{(2)}\nu^i}{Ds} = -\,k_1\nu^i + k_2\,{}_{(3)}\nu^i.$$

Hence $N_\nu(D\mu^i/Ds) = (dk_1/ds)\,{}_{(2)}\nu^i + k_1k_2\,{}_{(3)}\nu^i$ and an analogous equation obtains for $\bar{N}_\nu(\bar{D}\bar{\mu}^i/\bar{D}\bar{s})$. In virtue of these equations, (4.16) becomes

$$\frac{d\bar{k}_1}{d\bar{s}}\,{}_{(2)}\bar{\nu}^i + \bar{k}_1\bar{k}_2\,{}_{(3)}\bar{\nu}^i = e^{-3\sigma}\left[\frac{dk_1}{ds}\,{}_{(2)}\nu^i + k_1k_2\,{}_{(3)}\nu^i - \sigma_{hk}\nu^k(g^{ih} - \nu^i\nu^h)\right]. \tag{4.18}$$

We now find an equivalent expression for σ_{hk} which will permit us to write (4.18) in an invariant form. Under the assumption (4.15), the Riemann curvature tensors R_{hijk} of V_n and $\bar{R}_{hijk}$ of $\bar{V}_n$ exist and are continuous. The Ricci tensor R_{ij} and the invariant curvature R of V_n are defined by $R_{ij}=g^{hk}R_{hijk}$,

$R = g^{ij}R_{ij}$ with analogous definitions for the corresponding tensors in $\bar{V}_n$. A straightforward calculation, using (2.6) and (3.1) as well as the definitions of R_{hijk}, R_{ij} and R gives[48]

$$\begin{aligned} e^{-2\sigma}\bar{R}_{hijk} &= R_{hijk} + g_{hk}\sigma_{ij} + g_{ij}\sigma_{hk} - g_{hj}\sigma_{ik} - g_{ik}\sigma_{hj} \\ &\quad + (g_{hk}g_{ij} - g_{hj}g_{ik})\Delta_1\sigma, \end{aligned} \tag{4.19}$$

and

$$\begin{aligned} (n-1)(n-2)\sigma_{hk} &= (n-1)[\bar{R}_{hk} - R_{hk}] - \tfrac{1}{2}[\bar{g}_{hk}\bar{R} - g_{hk}R] \\ &\quad - \tfrac{1}{2}(n-1)(n-2)\Delta_1\sigma g_{hk} \end{aligned} \tag{4.20}$$

where $\Delta_1\sigma$ is the differential parameter of the first order defined by $\Delta_1\sigma = g^{ij}\sigma_{,i}\sigma_{,j}$. If[49] $n>2$, it follows from (2.6), (3.5) and (4.20) that (4.18) may be written as

$$\begin{aligned} &\frac{d\bar{k}_1}{d\bar{s}}\,{}_{(2)}\bar{\nu}^i + \bar{k}_1\bar{k}_2\,{}_{(3)}\bar{\nu}^i + \frac{1}{n-2}\bar{R}_{hk}\bar{\nu}^k(\bar{g}^{ih} - \bar{\nu}^i\bar{\nu}^h) \\ &\qquad = e^{-3\sigma}\left[\frac{dk_1}{ds}\,{}_{(2)}\nu^i + k_1k_2\,{}_{(3)}\nu^i + \frac{1}{n-2}R_{hk}\nu^k(g^{ih} - \nu^i\nu^h)\right]. \end{aligned}$$

This is the invariant form[50] of the law of transformation of $D\mu^i/Ds$.

We suppose that the members of this equation are not zero and write

$$J^2\eta^i = \frac{dk_1}{ds}\,{}_{(2)}\nu^i + k_1k_2\,{}_{(3)}\nu^i + \frac{1}{n-2}R_{hk}\nu^k(g^{ih} - \nu^i\nu^h) \tag{4.21}$$

---

(48) For example, cf. L. P. Eisenhart, loc. cit., pp. 89–90, especially equations (28.5) and (28.9).

(49) The assumption $n>2$ is to hold throughout §§5–12 inclusive. The case $n=2$ is considered in §14 and it is shown that the theorems which we obtain in §§5–12 apply in a modified sense.

(50) It is also possible to write the law of transformation of the principal normal given by (3.7) in an invariant form. For, as a consequence of (3.1),

$$n\sigma_{,j} = \left\{\overline{\begin{matrix} i \\ ij \end{matrix}}\right\} - \left\{\begin{matrix} i \\ ij \end{matrix}\right\}.$$

But

$$\left\{\begin{matrix} i \\ ij \end{matrix}\right\} = \frac{\partial}{\partial x^j}\log g^{1/2}$$

where $g = |g_{ik}|$ is the determinant whose elements are the components of the metric tensor. (For a proof of this statement, cf. L. P. Eisenhart, loc. cit., p. 18.) Hence

$$\sigma_{,j} = \frac{\partial}{\partial x^j}\log \bar{g}^{1/2n} - \frac{\partial}{\partial x^j}\log g^{1/2n}.$$

If we substitute this expression for $\sigma_{,j}$ in (3.7), it becomes

$$\bar{\mu}_i + \frac{\partial}{\partial x^k}\log \bar{g}^{1/2n}(\delta_i^k - \bar{\nu}^k\bar{\nu}_i) = \mu_i + \frac{\partial}{\partial x^k}\log g^{1/2n}(\delta_i^k - \nu^k\nu_i).$$

Thus a conformal geometric object is defined. As it is not a tensor, a formal theory based upon it would be quite complicated and we shall therefore not consider it further in this paper.

where $g_{ij}\eta^i\eta^j=1$ and write an analogous equation for $\bar{J}^2\bar{\eta}^i$. It follows that η^i is a unit conformetric vector, that is, the direction of η^i remains unchanged if V_n is subjected to a conformal transformation. We call η^i the *first conformal normal* of the curve C and shall also write it as ${}_{(1)}\eta^i$. The quantity J which is the unique positive root of J^2 is a relative conformal scalar having the transformation law

$$\bar{J} = e^{-\sigma}J. \tag{4.22}$$

We call J the *relative conformal curvature* of C. As a result of (2.4) and (4.22), S defined (up to an additive constant and choice of sign) by

$$S = \int J ds \tag{4.23}$$

remains invariant under conformal transformations and is a conformal scalar. We call S a *conformal arc length parameter* of C. If the values of the same arc length parameter are S_1 and S_2 at two points P_1 and P_2 of C, we call $|S_1-S_2|$ the *conformal arc length* or *conformal length* of the arc P_1P_2 (or P_2P_1) of C. It is clear that the conformal length is independent of the choice of the conformal arc length parameter. Since S is a conformal scalar, *the conformal length of an arc of a curve is unchanged by any conformal transformation of* V_n. The vector η^i and the parameter S have roles in the conformal theory analogous to those of ν^i and s in the metric theory.

As a consequence of (4.15) and (4.21), the parameter transformation (4.23) is of class C^1 so that we may apply the preceding results of this section with the conformal parameter S replacing t. We note that the inequalities

$$m \geqq p \geqq n+2 \tag{4.24}$$

which are analogous to and here replace (4.1) are sufficient to insure the existence of conformal Frenet equations for the vector η^i and the parameter S. The conditions (4.24) include (4.15). We summarize some of these remarks in the theorem

THEOREM 4.1. *Let C be a curve of class C^p $(p \geqq 3)$ in a V_n of class C^m $(m \geqq p)$ and dimensionality $n>2$ and let the relative conformal curvature of C be different from zero. Let the classes*[51] *of the first conformal normal ${}_{(1)}\eta^i$ and of each successive unit normal ${}_{(2)}\eta^i, {}_{(3)}\eta^i, \cdots, {}_{(r)}\eta^i$ obtained from ${}_{(1)}\eta^i$ by the Frenet process be C^1. Then there exists a set of scalars $J_1, J_2, \cdots, J_{r-1}$ $(r \leqq n-1)$ such that*

$$\frac{\delta\, {}_{(\alpha)}\eta^i}{\delta S} = -J_{\alpha-1}\, {}_{(\alpha-1)}\eta^i + J_\alpha\, {}_{(\alpha+1)}\eta^i, \quad J_0 = J_r = 0;\ \alpha = 1, 2, \cdots, r, \tag{4.25}$$

where S is a conformal arc length parameter. The ${}_{(\alpha)}\eta^i$ form a normalized r-bein

[51] It is unnecessary to assume the existence of the ${}_{(\alpha)}\eta^i$. For ${}_{(1)}\eta^i$ exists and the existence of ${}_{(\alpha)}\eta^i$ if $J_{\alpha-1}\neq 0$ follows from the fact that ${}_{(\alpha-1)}\eta^i$ has a derivative.

orthogonal to the tangent vector ν^i

$$g_{ij}\,{}_{(\alpha)}\eta^i\,{}_{(\beta)}\eta^j = \delta^\alpha_\beta, \qquad \alpha, \beta = 1, 2, \cdots, r,$$

$$g_{ij}\,{}_{(\alpha)}\eta^i\,\nu^j = 0.$$

If $C\leftrightarrow\overline{C}$, $V_n\leftrightarrow\overline{V}_n$ *by a conformal transformation, then equations analogous to* (4.25) *hold in* $\overline{V}_n$ *and at corresponding points of* C *and* $\overline{C}$, *the directions of the* ${}_{(\alpha)}\eta^i$ *correspond*

$$(4.26) \qquad {}_{(\alpha)}\bar{\eta}^i = e^{-\sigma}\,{}_{(\alpha)}\eta^i \qquad \alpha = 1, 2, \cdots, r,$$

and the J's are equal

$$\overline{J}_\gamma = J_\gamma, \qquad \gamma = 1, 2, \cdots, r-1.$$

We call (4.25) the *conformal Frenet equations*. In writing these equations we have replaced ${}_{(1)}\theta^i$, ${}_{(2)}\theta^i, \cdots, {}_{(r)}\theta^i$; $H_1, H_2, \cdots, H_{r-1}$ in (4.8), (4.9) and (4.12) by ${}_{(1)}\eta^i\,(=\eta^i)$, ${}_{(2)}\eta^i, \cdots, {}_{(r)}\eta^i$; $J_1, J_2, \cdots, J_{r-1}$ which we call the *first, second,* $\cdots$, *rth conformal normals; first, second,* $\cdots$, $(r-1)$*st conformal curvatures.*

The vector ${}_{(1)}\lambda^i$ defined by

$$(4.27) \qquad {}_{(1)}\lambda^i = J^{-1}\,{}_{(1)}\eta^i$$

is a conformal vector as follows from (4.22) and (4.26). Systems of conformal Frenet equations which involve only conformal vectors and conformal scalars may be derived corresponding to the initial conformal vector ${}_{(1)}\lambda^i$ by use of the derivatives defined in (3.24) and (3.26). We suppose that the relative conformal scalar Q defined in (3.25) is J. Then one easily finds as a consequence of (3.26), (3.28), (4.25) and (4.27) that

$$(4.28) \qquad \frac{\delta\,{}_{(\alpha)}\lambda^i}{\delta S_J} = -J_{\alpha-1}\,{}_{(\alpha-1)}\lambda^i + J_\alpha\,{}_{(\alpha+1)}\lambda^i, \quad J_0 = J_r = 0;\ \alpha = 1, 2, \cdots, r,$$

$$(4.29) \qquad \frac{\mathfrak{D}\,{}_{(\alpha)}\lambda^i}{\mathfrak{D}S} = -J_{\alpha-1}\,{}_{(\alpha-1)}\lambda^i + J_\alpha\,{}_{(\alpha+1)}\lambda^i, \quad J_0 = J_r = 0;\ \alpha = 1, 2, \cdots, r$$

where the conformal vectors ${}_{(\alpha)}\lambda^i$ correspond to the ${}_{(\alpha)}\eta^i$; that is,

$${}_{(\alpha)}\lambda^i = J^{-1}\,{}_{(\alpha)}\eta^i, \qquad {}_{(\alpha)}\bar{\lambda}^i = {}_{(\alpha)}\lambda^i.$$

5. **The $(n-1)$st conformal curvature** J_{n-1}. As a consequence of the remarks in the paragraph preceding (4.13), since η^i is orthogonal to ν^i, a curve can have at most $n-2$ nonzero conformal curvatures arising in the conformal Frenet equations. We shall now construct still another conformal invariant, unrelated to $J_1, J_2, \cdots, J_{r-1}$ and the conformal Frenet equations, which plays the role of the $(n-1)$st conformal curvature J_{n-1}. This curvature is as essential as the others in the development of the theory. However, the defini-

tion of J_{n-1} differs completely from that of the other conformal curvatures. Indeed, as we shall show later, there are important qualitative differences in the properties of the first $(n-2)$ conformal curvatures and the $(n-1)$st conformal curvature. This is natural because $J_1, J_2, \cdots, J_{n-2}$ are measures of the variation of the conformal normals whereas J_{n-1} is defined in connection with the variation of the relative conformal curvature.

It is assumed that (4.15) holds throughout this section except where it is explicitly replaced by another assumption. If we multiply the respective members of (3.6) and (3.7) by each other and sum,

$$\bar{k}_1^2 = \psi^2 k_1^2 + 2\psi\psi_{,i}\mu^i + \Delta_1\psi - \left(\frac{d\psi}{ds}\right)^2 \tag{5.1}$$

where $\Delta_1\psi = g^{ij}\psi_{,i}\psi_{,j}$ and

$$\psi = e^{-\sigma} \tag{5.2}$$

and $d\psi/ds = \psi_{,h}\nu^h$. We differentiate the last equation with respect to s and the form (2.1) and use (3.4). This gives

$$\frac{d^2\psi}{ds^2} = \psi_{,hk}\nu^h\nu^k + \psi_{,i}\mu^i.$$

When the value of $\psi_{,i}\mu^i$ from this equation is substituted in (5.1), it becomes

$$\bar{k}_1^2 = \psi^2 k_1^2 + 2\psi\frac{d^2\psi}{ds^2} - \left(\frac{d\psi}{ds}\right)^2 - 2\psi\psi_{,hk}\nu^h\nu^k + \Delta_1\psi. \tag{5.3}$$

Since, in virtue of (4.17) and (5.2),

$$\sigma_{hk} = -\psi_{,hk}/\psi,$$

(4.20) becomes

$$-2\psi\psi_{,hk} + \Delta_1\psi\cdot g_{hk}$$
$$= 2\psi^2\left[\frac{1}{n-2}(\bar{R}_{hk} - R_{hk}) - \frac{1}{2(n-1)(n-2)}(\bar{g}_{hk}\bar{R} - g_{hk}R)\right].$$

This equation and the fact that $\bar{\nu}^i = \psi\nu^i$ show that (5.3) is equivalent to

$$\bar{k}_1^2 + \bar{K} = \psi^2(k_1^2 + K) + 2\psi\frac{d^2\psi}{ds^2} - \left(\frac{d\psi}{ds}\right)^2, \tag{5.4}$$

where

$$K = \frac{1}{(n-1)(n-2)}\left[R - 2(n-1)R_{hk}\nu^h\nu^k\right], \tag{5.5}$$

and $\bar{K}$ is defined by an expression constructed from the analogous quantities

in V_n. If V_n is an S_n of constant curvature K, then $\mathcal{K}$ is found to be equal to K. If V_n is an Einstein space E_n of constant mean curvature ρ, then $\mathcal{K}=\rho/(n-1)$.

Now let $Q(s)$ be any nonzero relative conformal scalar of class C^2 which transforms according to the law

$$\bar{Q}(\bar{s}) = \psi Q(s). \tag{5.6}$$

According to (2.4) and (5.2), $d\bar{s}/ds=\psi^{-1}$. If we use this fact and differentiate (5.6) with respect to s, we obtain

$$\frac{d\psi}{ds}=\left[\frac{d\bar{Q}}{d\bar{s}}\cdot Q^2-\frac{dQ}{ds}\cdot\bar{Q}^2\right]/Q^2\bar{Q},$$

$$\psi\frac{d^2\psi}{ds^2}=\left[\frac{d^2\bar{Q}}{d\bar{s}^2}\bar{Q}Q^4-\frac{d^2Q}{ds^2}Q\bar{Q}^4-\frac{d\bar{Q}}{d\bar{s}}\frac{dQ}{ds}\bar{Q}^2Q^2\right.$$
$$\left.+2\left(\frac{dQ}{ds}\right)^2\bar{Q}^4-\left(\frac{d\bar{Q}}{d\bar{s}}\right)^2Q^4\right]/Q^4\bar{Q}^2.$$

Substitution of these results in (5.4) shows that

$$\left[2\bar{Q}\frac{d^2\bar{Q}}{d\bar{s}^2}-3\left(\frac{d\bar{Q}}{d\bar{s}}\right)^2-(\bar{k}_1^2+\bar{\mathcal{K}})\bar{Q}^2\right]/\bar{Q}^4$$
$$=\left[2Q\frac{d^2Q}{ds^2}-3\left(\frac{dQ}{ds}\right)^2-(k_1^2+\mathcal{K})Q^2\right]/Q^4.$$

This expression is equivalent to (5.1) if $Q(s)$ obeys (5.6). It is therefore the invariant form of the law of change of the first curvature of a curve when V_n is subjected to a conformal transformation. We summarize these results in the theorem

THEOREM 5.1. *Let $Q(s)$ be a nonzero relative conformal scalar defined along a curve in a V_n $(n>2)$ whose law of transformation is $\bar{Q}(\bar{s})=e^{-\sigma}Q(s)$. Then*

$$\left[2Q\frac{d^2Q}{ds^2}-3\left(\frac{dQ}{ds}\right)^2-(k_1^2+\mathcal{K})Q^2\right]/Q^4 \tag{5.7}$$

is a conformal scalar.

If we replace the assumption (4.15) by the stronger inequalities

$$m \geqq p \geqq 5, \tag{5.8}$$

then these conditions are sufficient that $J(s)$ defined by (4.21) and having the transformation law (4.22) be a relative conformal scalar of class C^2. Theorem 5.1 applied to $J(s)$ yields the conformal invariant

$$J_{n-1}=\left[2J\frac{d^2J}{ds^2}-3\left(\frac{dJ}{ds}\right)^2-(k_1^2+\mathcal{K})J^2\right]/J^4. \tag{5.9}$$

We call J_{n-1} the $(n-1)$*st conformal curvature* of the curve C. If J is referred to a conformal arc length parameter S, we find by use of (4.23) that (5.9) becomes

$$J_{n-1} = \left[2J \frac{d^2J}{dS^2} - 3\left(\frac{dJ}{dS}\right)^2 - (k_1^2 + K)\right]/J^2. \tag{5.10}$$

In a similar way, the conformal scalar (5.7) may be written in terms of Q and its derivatives with respect to S.

We note that if J_{n-1} exists in a V_n of class C^3 then it follows from the definition of J_{n-1} and (4.22) that $\bar{J}_{n-1}$ must exist in $\bar{V}_n$. This observation plus a similar one implicit in the italicized statement containing (4.12) shows that *if the conformal curvatures $J_1, J_2, \cdots, J_{r-1}, J_{n-1}$ exist for a curve C in a V_n of class C^3, then the conformal curvatures also exist for the conformal image curve $\bar{C}$ in $\bar{V}_n$*. Since the conformal arc length of a curve remains invariant under conformal transformations, it is always possible to pick conformal arc length parameters S and $\bar{S}$ on C and $\bar{C}$ so that corresponding points are given by $S=\bar{S}$. Then, according to Theorem 4.1 and the preceding paragraphs, all the conformal curvatures (if they exist) and S are the same at corresponding points of C and $\bar{C}$. As an immediate consequence of this remark, we have the theorem

THEOREM 5.2. *If $n>2$ and $V_n \leftrightarrow \bar{V}_n$, $C \leftrightarrow \bar{C}$ by a conformal map, then a conformal arc length parameter S may be chosen so that corresponding points of C and $\bar{C}$ have the same value of S and the conformal curvatures $J_1, J_2, \cdots, J_{n-1}$ are the same functions of S for C and $\bar{C}$.*

The analogous theorem in the metric theory of a curve is well known.

6. **The existence theorem.** In §§4 and 5, it was shown that any curve in V_n which satisfies certain general conditions determines (except for sign) r $(\leqq n-1)$ continuous conformal curvatures. We now prove the following converse:

THEOREM 6.1. *Let V_n be any Riemann space of class C^4 and dimensionality $n>2$. Suppose that*

$$J_1(S), J_2(S), \cdots, J_{r-1}(S), J_{n-1}(S), \qquad r \leqq n-1; a < S < b, \tag{6.1}$$

are any continuous functions no one of which, except possibly J_{n-1} vanishes identically in any subinterval of $a<S<b$. Let x_0^i be the coordinates of any point P of V_n; J_0, L_0 any two real numbers of which the first is positive; v_0^i, ${}_{(1)}\eta_0^i$, ${}_{(2)}\eta_0^i$, $\cdots$, ${}_{(r)}\eta_0^i$ any normalized $(r+1)$-bein at P; μ_0^i any vector at P orthogonal to v_0^i; and S_0 any real number so that $a<S_0<b$. Then there exists a curve

$$x^i = x^i(S) \tag{6.2}$$

of class C^3 defined in some subinterval of $a<S<b$ about S_0 which has S as a

conformal arc length parameter and the functions (6.1) *as conformal curvatures. For* $S=S_0$, *the curve passes through* P *so that its moving conformal* $(\tau+1)$-*bein and principal normal take the positions* ν_0^i, ${}_{(1)}\eta_0^i$, ${}_{(2)}\eta_0^i, \cdots, {}_{(\tau)}\eta_0^i$ *and* μ_0^i *respectively and the values of the relative conformal curvature and its derivative with respect to* S *at* P *are* J_0 *and* L_0 *respectively. Any other curve with these properties will coincide with* (6.2) *in the common interval of definition.*

Briefly, this theorem states that curves exist in V_n whose conformal curvatures are any arbitrary continuous functions of the conformal arc length and that any such curve is uniquely determined by a number of initial conditions. For the proof, we consider the system of $[2+n(\tau+3)]$ differential equations

$$
\begin{aligned}
&\frac{dx^i}{dS} = J^{-1}\nu^i,\\
&\frac{d\nu^i}{dS} = J^{-1}\left(-\begin{Bmatrix} i \\ jk \end{Bmatrix}\nu^j\nu^k + \mu^i\right),\\
&\frac{d\mu^i}{dS} = J^{-1}\left[-\begin{Bmatrix} i \\ jk \end{Bmatrix}\mu^j\nu^k - g_{jk}\mu^j\mu^k\nu^i + J^2\,{}_{(1)}\eta^i\right.\\
&\qquad\qquad\left. - \frac{1}{n-2}R_{jk}\nu^j(g^{ki}-\nu^k\nu^i)\right],\\
&\frac{d\,{}_{(1)}\eta^i}{dS} = -J^{-1}\begin{Bmatrix} i \\ jk \end{Bmatrix}{}_{(1)}\eta^j\nu^k - J^{-1}g_{jk}\,{}_{(1)}\eta^j\mu^k\nu^i + J_1(S)\cdot{}_{(2)}\eta^i,\\
(6.3)\quad &\frac{d\,{}_{(2)}\eta^i}{dS} = -J^{-1}\begin{Bmatrix} i \\ jk \end{Bmatrix}{}_{(2)}\eta^j\nu^k - J^{-1}g_{jk}\,{}_{(2)}\eta^j\mu^k\nu^i - J_1(S)\,{}_{(1)}\eta^i + J_2(S)\,{}_{(3)}\eta^i,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,\\
&\frac{d\,{}_{(\tau)}\eta^i}{dS} = -J^{-1}\begin{Bmatrix} i \\ jk \end{Bmatrix}{}_{(\tau)}\eta^j\nu^k - J^{-1}g_{jk}\,{}_{(\tau)}\eta^j\mu^k\nu^i - J_{\tau-1}(S)\,{}_{(\tau-1)}\eta^i,\\
&\frac{dJ}{dS} = L,\\
&\frac{dL}{dS} = \frac{J^{-1}}{2}\left[J^2J_{n-1}(S) + 3L^2 + g_{jk}\mu^j\mu^k\right.\\
&\qquad\qquad\left. + \frac{1}{(n-1)(n-2)}\{R - 2(n-1)R_{jk}\nu^j\nu^k\}\right]
\end{aligned}
$$

in the $[2+n(\tau+3)]$ unknowns J, L, x^i, ν^i, μ^i ${}_{(1)}\eta^i$, ${}_{(2)}\eta^i, \cdots, {}_{(\tau)}\eta^i$. This system of equations has been obtained by rewriting equations (3.3), (3.4), (4.21), (4.25) and (5.10) in the normal form for ordinary differential equations using the definitions of the conformal derivative, S and $\mathcal{K}$ which are given by (3.11), (4.22) and (5.5) respectively. The right members are continuous for

any x^i in the coordinate system; any ν^i, μ^i, ${}_{(\alpha)}\eta^i$ $(\alpha=1, 2, \cdots, \tau)$; any J different from zero; any L; and any S in the interval $a<S<b$. They are of class C^1 in all the dependent variables since the g_{ij} are of class C^3. Therefore we can apply the fundamental existence theorem for such a system of differential equations.

According to this theorem, there exists a set of functions $J(S)$, $L(S)$, $x^i(S)$, $\nu^i(S)$, $\mu^i(S)$, ${}_{(\alpha)}\eta^i(S)$ of class C^1 satisfying (6.3) which are defined in a sufficiently small interval I of $a<S<b$ about S_0 and which assume the values J_0, L_0, ν_0^i, μ_0^i, ${}_{(\alpha)}\eta_0^i$ for $S=S_0$. Any two solutions of (6.3) which have the same initial conditions are identical in their common interval of definition. It is readily seen that $J(S)$ is of class C^2. Hence, as a consequence of the special form of the first two equations, $x^i(S)$ are actually of class C^3. The existence theorem applies to a fixed coordinate system. Nevertheless the solutions in one coordinate system will transform under a change of coordinates into the solutions in any other since the differential equations (6.3) may be written in the invariant form, tensor = tensor.

The solutions $x^i(S)$ determine a curve C whose parametric equations are (6.2). We now show that the dependent variables in (6.3) as well as the J's actually have the geometric significance for C that is stated in the theorem. Let

$$\begin{aligned} A_{\alpha\beta}(S) &= g_{jk}\, {}_{(\alpha)}\eta^j(S)\, {}_{(\beta)}\eta^k(S), \\ A_{00}(S) &= g_{jk}\, \nu^j(S)\nu^k(S), \\ A_{0\alpha}(S) &= g_{jk}\nu^j(S)\, {}_{(\alpha)}\eta^k(S), \\ B(S) &= g_{jk}\nu^j(S)\mu^k(S), \\ C_\alpha(S) &= g_{jk}\mu^j(S)\, {}_{(\alpha)}\eta^k(S). \end{aligned}$$

We prove that these quantities keep their initial values, that is,

$$\begin{aligned} A_{\alpha\beta}(S) &= \delta_\beta^\alpha, & A_{00}(S) &= 1, \\ A_{0\alpha}(S) &= 0, & B(S) &= 0 \end{aligned} \tag{6.4}$$

all along C. A straightforward calculation using (6.3) shows that

$$\begin{aligned} \frac{dA_{\alpha\beta}}{dS} &= -J^{-1}C_\alpha A_{0\beta} - J_{\alpha-1}A_{(\alpha-1)\beta} + J_\alpha A_{(\alpha+1)\beta} \\ &\quad - J^{-1}C_\beta A_{0\alpha} - J_{\beta-1}A_{(\beta-1)\alpha} + J_\beta A_{(\beta+1)\alpha}, \\ \frac{dA_{00}}{dS} &= 2BJ^{-1}, \\ \frac{dA_{0\alpha}}{dS} &= J^{-1}C_\alpha(1 - A_{00}) - J_{\alpha-1}A_{0(\alpha-1)} + J_\alpha A_{0(\alpha+1)}, \\ \frac{dB}{dS} &= J^{-1}g_{jk}\mu^j\mu^k(1 - A_{00}) + JA_{01}. \end{aligned} \tag{6.5}$$

The derivation of these equations is somewhat simplified by noting that $dA_{\alpha\beta}/dS = DA_{\alpha\beta}/DS$, and so on.

In these equations, we think of $A_{\alpha\beta}$, A_{00}, $A_{0\alpha}$, B as independent variables and the other quantities as known functions of S given by the solution of (6.3). Then (6.5) is a system of ordinary differential equations in the normal form, and the right members satisfy the conditions for the existence of a unique solution determined by arbitrary initial values of the A's and B and for S in the interval I on which the solution of (6.3) is defined. It is easy to verify that (6.4) is a solution of (6.5) for S in I. Since the ν_0^i, ${}_{(\alpha)}\eta_0^i$, μ_0^i were chosen so that (6.4) holds for $S = S_0$, it follows that (6.4) is true for any S belonging to I. Consequently ν^i is orthogonal to μ^i and ν^i, ${}_{(\alpha)}\eta^i$ form a normalized $(\tau+1)$-bein all along C. Since $A_{00} = 1$, $\nu^i = dx^i/ds$ where s is a metric arc length parameter and $dS = Jds$. It is then readily seen from the second equation in (6.3) that μ^i is the principal (metric) normal, and from the third equation that J and ${}_{(1)}\eta^i$ are the relative conformal curvature and the first conformal normal of C. After this result, it is immediate that S is a conformal arc length parameter, the J's are the conformal curvatures and the ${}_{(\alpha)}\eta^i$ are the conformal normals of C. This completes the proof of the theorem.

The set of geometric objects $M = \{x^i, \nu^i, {}_{(\alpha)}\eta^i, \mu^i, J, L\}$ which together with S comprise the set of initial conditions of the theorem may be defined independently of any curve in V_n. The ν^i, ${}_{(\alpha)}\eta^i$ form an arbitrary normalized $(\tau+1)$-bein and μ^i is any vector normal to ν^i and J and L are any scalars of which the first is positive. Under any conformal transformation of V_n, the respective objects of M transform according to the laws (2.5), (3.5), (4.26), (3.7), (4.22) and

$$\bar{L} = e^{-\sigma}(L - \sigma_{,j}\nu^j) \tag{6.6}$$

respectively. The set $\{x^i, \nu^i, {}_{(\alpha)}\eta^i, J, L\}$ is called an *M-set*. The *order* of the M-set is the integer $\tau+1$ whose maximum value is n. The role of an M-set of order $\tau+1$ in the conformal theory of a curve is analogous to that of a normalized $(\tau+1)$-bein in the metric theory. If the geometric objects of an M-set have the geometric significance described in Theorem 6.1 for some curve C, then the M-set is said to be *associated* with C at the point whose coordinates belong to M.

One may refer the curve C to a metric arc length parameter s. For (4.23) defines S as a function of s which may be substituted in the equations (6.2) of C. Since J has two continuous derivatives and $x^i(S)$ are of class C^3, C is also of class C^3 when s is the parameter of the curve.

If V_n is an Einstein space E_n of constant mean curvature ρ then since

$$R_{ij} = -\rho g_{ij} \tag{6.7}$$

the term $R_{jk}\nu^k(g^{ij} - \nu^i\nu^j)/(n-2)$ is identically zero in the third equation of (6.3). As a result, it is readily seen that in this case the hypothesis of Theo-

rem 6.1 may be weakened so that E_n is a space of class C^3 instead of class C^4 and the remainder of the theorem will hold as stated.

We now show that if V_n is an $\overline{E}_n$, the hypothesis of Theorem 6.1 may be weakened in the same manner as for an E_n. Suppose that $\overline{E}_n$ is of class C^3 and that $\sigma(x^i)$ is a mapping function (necessarily of class C^2) which maps $\overline{E}_n$ conformally on E_n. Let $J_1(S)$, $J_2(S)$, $\cdots$, $J_{r-1}(S)$, $J_{n-1}(S)$ be the preassigned conformal curvatures and let an M-set $\overline{M}\equiv\{\bar{x}_0^i, \bar{\nu}_0^i, {}_{(\alpha)}\bar{\eta}_0^i, \bar{\mu}_0^i, \overline{J}_0, \overline{L}_0\}$ and S_0 be a set of initial conditions of the sort enumerated in Theorem 6.1 and suppose that M, S_0 is the set of quantities in E_n corresponding to these initial conditions under the mapping determined by $\sigma(x^i)$. In accordance with the preceding paragraph the set M, S_0 and the J's determine a curve C of class C^3 in E_n. Geometric objects like those of the set M exist at all points of C and have continuous first derivatives.

Let $\overline{C}$ be the image of C in $\overline{E}_n$. Since $\overline{C}$ is of class C^3, $\bar{\nu}^i$, $\bar{\mu}^i$, ${}_{(1)}\bar{\eta}^i$, $\overline{J}$ all exist. As a result of the existence of $\overline{J}$, (4.22) holds. From (4.22) and the fact that J and σ are each of class C^2, it follows that $\overline{J}$ is also so that $\overline{L}$ exists and is of class C^1. It follows that S is a conformal arc length parameter for $\overline{C}$ as well as for C. Both curves have the same parametric equations in terms of the parameter S.

Since ${}_{(1)}\bar{\eta}^i$ exists, ${}_{(1)}\bar{\eta}^i=e^{-\sigma}{}_{(1)}\eta^i$. From the classes of σ and ${}_{(1)}\eta^i$, we infer that ${}_{(1)}\bar{\eta}^i$ has a continuous derivative so that the first conformal Frenet equation

$$\frac{\bar{\delta}\ {}_{(1)}\bar{\eta}^i}{\bar{\delta} S}=\overline{J}_1\ {}_{(2)}\bar{\eta}^i$$

holds. It follows from Theorem 4.1 that $\overline{J}_1=J_1$ and ${}_{(2)}\bar{\eta}^i=e^{-\sigma}{}_{(2)}\eta^i$. Hence ${}_{(2)}\bar{\eta}^i$ has a continuous first derivative and we may proceed as before. In this way one shows that all the conformal normals exist and have continuous first derivatives and that the (existent) conformal curvatures of $\overline{C}$ are the preassigned functions J_1, J_2, $\cdots$, J_{r-1}, J_{n-1}. The initial conditions given by $\overline{M}$, S_0 must be satisfied because of the manner in which C was constructed. The unique determination of $\overline{C}$ follows readily from the known uniqueness of C. Thus the conclusions of Theorem 6.1 hold in the case of an $\overline{E}_n$ even if the class of $\overline{E}_n$ is only C^3. In particular, the existence theorem applies when the space is an $\overline{R}_n$ of class C^3.

7. **The conformal equivalence theorem.** The fundamental theorem in the metric theory is the congruence theorem for a curve in a euclidean space or a space of constant curvature: All curves in S_n with equal curvatures $k_\alpha(s)$ are congruent, that is, they may be made to coincide by a motion in S_n. We now develop some minor results leading to the analogous theorem in the conformal theory of a curve in any conformally euclidean space $\overline{R}_n$. The present proof is based upon the existence theorem. Another proof which does not use the results of Theorem 6.1 exists and depends upon the following ideas: Suppose

C_1 and C_2 are curves in two conformally euclidean spaces whose conformal curvatures are the same functions of their conformal arc lengths. Then a mapping is established between C_1 and C_2 so that points with equal values of the conformal arc length parameters correspond. It may then be shown that this mapping may be imbedded in a conformal transformation between the two spaces. This conformal transformation necessarily maps the spaces on each other in such a manner that C_1 and C_2 correspond. The proof just outlined will not be given in this paper.

If an S_n of constant curvature K is mapped conformally on an S_n' of constant curvature K' where S_n and S_n' are spaces of class C^3 and $n>2$, then one readily finds from (4.19) and (4.20) that the mapping function $\sigma(x^i)$ satisfies the differential equations

$$\sigma_{ij} = -\tfrac{1}{2}\left[e^{2\sigma}K' - K + \Delta_1\sigma\right]g_{ij}. \tag{7.1}$$

Conversely, if a transformation whose mapping function is a solution of (7.1) is applied to S_n, the image space must be S_n'. Suppose that both S_n and S_n' are euclidean spaces. Then $K=K'=0$. In this case (7.1) may be written as

$$\psi_{,ij} = \frac{\Delta_1\psi}{2\psi}\cdot g_{ij} \tag{7.2}$$

where ψ is defined by (5.2). Let the x^i be cartesian rectangular coordinates. Then $g_{ij}=\delta^i_j$ and (7.2) becomes

$$\frac{\partial^2\psi}{\partial x^i \partial x^j} = \frac{\Delta_1\psi}{2\psi}\delta^i_j.$$

The solution of these equations is easily found to be

$$\psi = a, \qquad a > 0, \tag{7.3}$$

or

$$\psi = \sum_{i=1}^{n} b(x^i - d^i)^2, \qquad b > 0, \tag{7.4}$$

where a, b, d^i are real constants. The point given by $x^i=d^i$ is a singular point of any conformal transformation associated with (7.4). It follows that *every mapping function that maps R_n $(n>2)$ on itself conformally must satisfy* (7.3) *or* (7.4) *and conversely*. We consider the group of conformal transformations of R_n into itself more fully in the next section. At present, we only prove the following theorem:

THEOREM 7.1. *A conformal transformation of R_n $(n>2)$ into itself exists which transforms any given M-set M into any other given M-set $\overline{M}$ of the same order.*

It is shown in §8 that if the order of M is n then the transformation is uniquely determined. The proof of the theorem follows. Let the sets M and $\overline{M}$ be $\{x^i, \nu^i, {}_{(\alpha)}\eta^i, \mu_i, J, L\}$ and $\{\bar{x}^i, \bar{\nu}^i, {}_{(\alpha)}\bar{\eta}^i, \bar{\mu}_i, \bar{J}, \bar{L}\}$ respectively. We consider two cases.

(1) Suppose $\bar{\mu}_i = \mu_i$ and $\bar{L}/\bar{J} = L/J$ at the point P whose coordinates x^i belong to M. Then a unique positive([82]) constant a exists so that $\bar{J} = aJ$, $\bar{L} = aL$. If we subject R_n to the magnification of similitude with center at P determined by $\psi = e^{-\sigma} = a$ it follows from (3.7), (4.22) and (6.6) that $\{\mu_i, J, L\}$ transform into $\{\bar{\mu}_i, \bar{J}, \bar{L}\}$. Suppose that the magnification transforms ν^i, ${}_{(\alpha)}\eta^i$ into ν'^i, ${}_{(\alpha)}\eta'^i$. Then the normalized bein ν'^i, ${}_{(\alpha)}\eta'^i$ at P may be mapped on $\bar{\nu}^i$, ${}_{(\alpha)}\bar{\eta}^i$ at $\bar{P}$ by a motion in R_n where $\bar{P}$ is the point whose coordinates $\bar{x}^i$ belong to $\overline{M}$. Since $\bar{\mu}_i, \bar{J}, \bar{L}$ are metric geometric objects, this motion leaves them unchanged. Hence M is transformed into $\overline{M}$ by means of a magnification of similitude (7.3) followed by a motion.

(2) Suppose at least one of the equations in the hypothesis of case (1) is untrue. We choose cartesian rectangular coordinates in R_n so that P is the origin of coordinates and $\nu^1 = \nu_1 = 1$, $\nu^\gamma = \nu_\gamma = 0$, $(\gamma = 2, 3, \cdots, n)$. It follows that $\mu_1 = \bar{\mu}_1 = 0$. We now determine unique constants b, d^i so that a conformal transformation associated with (7.4) transforms M into $\overline{M}$. At the origin, we find from (7.4) that

$$e^{-\sigma} = b\sum_{j=1}^{n} d^{j^2} \qquad \sigma_{,i} = 2d^i/\sum_j d^{j^2}.$$

If we substitute these values in (3.7), (4.22) and (6.6) and write

$$A_i = (\mu_i - \bar{\mu}_i)/2, \qquad B = \bar{J}/J, \qquad D = (\bar{J}L - J\bar{L})/2J$$

we find

$$A_1 = 0, \qquad d^\gamma = A_\gamma \sum_j d^{j^2}, \qquad \gamma = 2, 3, \cdots, n,$$

$$B = b\sum_j d^{j^2}, \qquad D = bd^1.$$

These equations lead to

$$bB = B^2\sum_{\gamma=2}^{n} A_\gamma^2 + D^2.$$

Since $J > 0$ and $\bar{J} > 0$, $B > 0$ and the above equation gives a unique solution for b. The value of b obviously cannot be negative and also cannot be zero since in this latter case $A_1 = A_2 = \cdots = A_n = D = 0$ which is impossible according to the hypothesis. Then $d^1 = D/b$ and $d^\gamma = BA_\gamma/b$. The unique numbers $d^1, d^2, \cdots, d^n$ cannot all be zero for the reason just given. Hence the origin is

([82]) The positive sign of a and, in case (2) below, of b is due to our definition of J as non-negative. If both positive and negative J's were permitted, the present discussion and the conformal equivalence theorem would be needlessly complicated.

not a singular point of any transformation associated with (7.4). Any such transformation T would transform $\{\mu_i, J, L\}$ into $\{\bar{\mu}_i, \bar{J}, \bar{L}\}$. If it also changes ν^i, ${}_{(\alpha)}\eta^i$ at P into ν'^i, ${}_{(\alpha)}\eta'^i$ at P' then the normalized bein ν'^i, ${}_{(\alpha)}\eta'^i$ at P' may be transformed into $\bar{\nu}^i$, ${}_{(\alpha)}\bar{\eta}^i$ at $\bar{P}$ by a motion T_1 of R_n. Then T_1T transforms M into $\bar{M}$. The proof of the theorem is thus completed. A similar theorem exists for any $\bar{R}_n$ and may be proved by mapping $\bar{R}_n$ conformally on R_n and using Theorem 7.1. We may now easily prove the fundamental conformal equivalence theorem:

THEOREM 7.2. *Let ${}_{(1)}\bar{R}_n$ and ${}_{(2)}\bar{R}_n$ be conformally euclidean spaces of class C^3 and dimensionality $n>2$ and let C_1 and C_2 be curves in ${}_{(1)}R_n$ and ${}_{(2)}R_n$ respectively whose conformal curvatures are the same functions of the conformal arc length. Then a conformal transformation exists so that ${}_{(1)}\bar{R}_n \leftrightarrow {}_{(2)}\bar{R}_n$ and $C_1 \leftrightarrow C_2$.*

In short, this theorem states that curves in $\bar{R}_n$'s whose $J_1, J_2, \cdots, J_{r-1}, J_{n-1}$ are the same functions of S are conformally equivalent. To prove the theorem, we note that conformal transformations T_1 and T_2 exist which map ${}_{(1)}\bar{R}_n$ and ${}_{(2)}\bar{R}_n$ respectively on R_n. For these transformations we also have $C_1 \rightarrow \bar{C}_1$ and $C_2 \rightarrow \bar{C}_2$ respectively where $\bar{C}_1$ and $\bar{C}_2$ are curves in R_n. Since $J_1, J_2, \cdots, J_{r-1}, J_{n-1}$ exist for C_1 and C_2, in accordance with the italicized statement preceding Theorem 5.2, they exist for $\bar{C}_1$ and $\bar{C}_2$ also. As a consequence of Theorem 5.2, the J's for $\bar{C}_1$ and $\bar{C}_2$ are the same functions of their respective conformal arc lengths S.

Let P_1 and P_2 be two points which belong to $\bar{C}_1$ and $\bar{C}_2$ respectively such that the conformal arc length parameters for $\bar{C}_1$ and $\bar{C}_2$ at P_1 and P_2 have the same value S_0. Let M_1 and M_2 be the M-sets associated with $\bar{C}_1$ and $\bar{C}_2$ respectively at P_1 and P_2. As a consequence of Theorem 7.1, a conformal transformation T of R_n into itself exists which transforms M_1 into M_2. This same transformation transforms $\bar{C}_1$ into some curve C' passing through P_2 for $S=S_0$ and having the associated M-set M_2 at P_2. For reasons like those given in connection with $\bar{C}_1$ and $\bar{C}_2$, the J's for C' and $\bar{C}_2$ are the same functions of S. It follows from Theorem 6.1 that C' coincides with $\bar{C}_2$ in a sufficiently small neighborhood of P_2. Hence the conformal transformation $T_2^{-1}TT_1$ transforms ${}_{(1)}\bar{R}_n$ into ${}_{(2)}\bar{R}_n$ mapping sufficiently small arcs of C_1 and C_2 on each other. This proves Theorem 7.2. As a result of Theorem 6.1 and Theorem 7.2, the equations

$$J_1 = J_1(S),\ J_2 = J_2(S),\ \cdots,\ J_{r-1} = J_{r-1}(S),\ J_{n-1} = J_{n-1}(S)$$

may be regarded as conformal intrinsic equations of a curve in $\bar{R}_n$ determining the curve up to a conformal transformation of the space. A detailed conformal geometry of curves could be developed by a study of important particular conformal intrinsic equations.

8. Groups of conformal transformations in euclidean space R_n and in a conformally euclidean space $\bar{R}_n$. A euclidean space R_n admits a group G of

conformal transformations of the space. According to the italicized statement preceding Theorem 7.1, the mapping function which is associated with any of these transformations must satisfy (7.3) or (7.4). Suppose that T_1 and T_2 are two transformations of G which are associated with the same ψ. Then T_1 and T_2 induce changes in the metric element of arc ds of R_n which may be written as

$$ds_1 = \psi^{-1}ds, \qquad ds_2 = \psi^{-1}ds$$

respectively. If we write the point transformations as

$$T_1(P) = P_1, \qquad T_2(P) = P_2$$

then the point transformation T_3 defined by $T_3(P_1) = P_2$ is a conformal transformation belonging to G for which the induced change in metric is $ds_1 = ds_2$. Hence T_3 is a euclidean motion and T_2 may be written as $T_2 = T_3T_1$.

If ψ is given by (7.3), one transformation associated with ψ is a magnification of similitude having any point of R_n as center. If ψ is defined by (7.4), one transformation associated with ψ is readily found to be the inversion with respect to the hypersphere whose center is given by $x^i = d^i$ and whose radius r is $b^{-1/2}$. For the equations of this inversion are

$$x'^i = \frac{r^2(x^i - d^i)}{\sum_j (x^j - d^j)^2} + d^i.$$

From these equations, we find that

$$\sum_i dx'^{i^2} = \frac{r^4}{[\sum_j (x^j - d^j)^2]^2} \cdot \sum_i dx^{i^2}$$

which is a conformal transformation of R_n with ψ given by (7.4). These results and the remarks of the first paragraph of this section prove the theorem of Liouville([53]).

THEOREM 8.1. *The most general conformal map of R_n $(n>2)$ on itself is the product of an inversion with respect to a hypersphere by a motion or the product of a magnification of similitude by a motion.*

Now in the proof of Theorem 7.1, it was noted that the geometric objects $\{\mu_i, J, L\}$ belonging to the two M-sets M and $\overline{M}$ uniquely determine the constants a or b and d^i which define a mapping function associated with a transformation of G. Hence if the orders of M and $\overline{M}$ are each n so that the ν^i, ${}_{(\alpha)}\eta^i$ form normalized n-biens, it is readily seen in consequence of Theorem 8.1 that the two M-sets determine a *unique* transformation of G.

([53]) For $n>3$, this theorem was proved by S. Lie, *Über Complexe, insbesondere Linien- und Kugelcomplexe, mit Anwendung auf die Theorie partieller Differentialgleichungen*, Mathematische Annalen, vol. 5 (1872), pp. 145–246. Also cf. L. Bianchi, *Lezioni di Geometria Differenziale*, 2d edition, vol. 1, 1902, pp. 375, 376.

Suppose that the geometric objects in $\overline{M}$ are fixed while those which belong to M range over all admissible values. Then the corresponding conformal transformations range over the totality of transformations belonging to G. Hence the geometric objects of M determine the parameters of G. The normalized n-bein at any point provides $n(n+1)/2$ independent constants, the vector μ^i orthogonal to ν^i contributes $n-1$ additional constants and J, L are two more parameters. As a result the group G has exactly $(n+1)(n+2)/2$ essential parameters.

Let $\overline{R}_n$ $(n>2)$ be a conformally euclidean space and suppose that T_1 is a conformal transformation which transforms $\overline{R}_n$ into R_n. If T is any transformation belonging to G, then the conformal transformation $\overline{T}=T_1^{-1}TT_1$ maps $\overline{R}_n$ on itself. The totality of these transformations $\{\overline{T}\}$ form the complete group $\overline{G}$ of conformal transformations of $\overline{R}_n$ upon itself. The group $\overline{G}$ is the conformal image in $\overline{R}_n$ of the group G in R_n and is obviously independent of the particular mapping T_1. In consequence of the preceding discussion we have the theorem.

Theorem 8.2. *Every conformally euclidean space $\overline{R}_n$ $(n>2)$ admits a continuous group of conformal transformations on itself having $(n+1)(n+2)/2$ essential parameters.*

Any path in R_n of the group G is a curve C which is described by a point as the latter undergoes the transformations of a one parameter subgroup of G. If P_1 and P_2 are any two points of C, a conformal transformation belonging to this subgroup exists which maps C on itself so that P_1 coincides with P_2. Now it may be shown that if[64] the relative conformal curvature $J\neq 0$ then the conformal curvatures $J_1, J_2, \cdots, J_{r-1}, J_{n-1}$ of C exist. Hence it follows from Theorem 5.2 that the J's have the same values at P_1 and P_2 and therefore each conformal curvature is constant along the curve.

Conversely, let C be a curve of R_n each of whose conformal curvatures is equal to a constant. In consequence of the conformal equivalence theorem, a transformation of G exists which maps C on itself so that any two given points P_1 and P_2 coincide. These transformations may be chosen (in those cases where they are not already determined) so that they belong to a one parameter subgroup of G. Hence C is a path of the group G.

The paths in $\overline{R}_n$ of the group $\overline{G}$ are readily seen to be the conformal images of the paths in R_n of G. If $\overline{R}_n$ is of class C^3, the italicized statement preceding Theorem 5.2 shows that the conformal curvatures of each path of $\overline{G}$ exist. If the line of reasoning employed in the discussion of the paths of G is now followed, we arrive at the following theorem:

Theorem 8.3. *The paths whose relative conformal curvatures do not vanish*

---

(64) It is easy to prove, using the results of §10, that every curve in $\overline{R}_n$ whose relative conformal curvature vanishes is a path of $\overline{G}$.

of the group of conformal transformations of an $\bar{R}_n$ $(n>2)$ *of class* C^3 *upon itself coincide with the curves in* $\bar{R}_n$ *each of whose conformal curvatures is equal to a constant.*

The results of this and the preceding sections obviously apply to a curve in R_n which is subjected to any transformation of the continuous group G. Consequently, at least for curves in R_n, the results of this paper might have been obtained by using the classical methods of Lie and his followers which depend upon the elimination of the parameters of G. Our results would thus constitute the inversive theory of curves when applied to the inversive group G. The present methods based upon the conformal derivative yield considerably simpler proofs than would be obtained by elimination of the parameters of G and the results have greater applicability, being valid for a curve in any V_n.

From the standpoint of Lie theory, this inversive geometry is simply the "natural geometry" of curves in an R_n under the transformations of G. It is analogous to the well known natural geometry of curves in an R_n associated with the metric group. It has been shown by Pick([55]) that it is possible to develop a "natural geometry" of curves (at least when the enveloping space is an R_3) for an arbitrary continuous group and this theory has been carried further by Kowalewski([56]) and his students using the methods of the classical Lie theory.

9. **Conformal differential invariants.** A conformal differential invariant H of a curve C in V_n whose equations are $x^i = x^i(t)$ is a scalar function of the variables:

$$(9.1)\qquad x^i(t),\ \frac{dx^i}{dt},\ \frac{d^2x^i}{dt^2},\ \cdots,\ g_{ij}[x^k(t)],\ \frac{\partial g_{ij}}{\partial x^h},\ \frac{\partial^2 g_{ij}}{\partial x^h \partial x^k},\ \cdots,$$

defined along C which, at any point P of C, has the same value for any admissible change of coordinates x^i or of the parameter t and whose value also remains unchanged if V_n is mapped conformally on $\bar{V}_n$. This last condition, which gives H its conformal character, is equivalent to the assumption that H is invariant if the g_{ij} and their derivatives are replaced respectively by $e^{2\sigma}g_{ij}$ and their derivatives while the other variables in (9.1) are unchanged.

The simplest conformal differential invariants of C are its conformal curvatures. The most important problem concerning conformal invariants is the discovery of all such invariants. Before answering this question (at least for curves in an $\bar{R}_n$) we discuss the relationship between two apparently distinct processes for constructing new conformal differential invariants from known invariants.

([55]) G. Pick, loc. cit., p. 139.
([56]) G. Kowalewski, loc. cit., chap. 3.

If H is a conformal scalar, then dH/dS, where S is a conformal arc length parameter, is also a conformal invariant. This classical method of constructing new invariants is a direct consequence of the properties of S. In particular, any function of the conformal curvatures and their derivatives with respect to S is a conformal invariant. A different method is apparently provided by Theorem 5.1. For the relative conformal scalar Q defined by

$$Q = J \cdot H \tag{9.2}$$

obeys (5.6) and it follows from Theorem 5.1 that the scalar

$$q = \left[2Q \frac{d^2Q}{ds^2} - 3\left(\frac{dQ}{ds}\right)^2 - (k_1^2 + K)Q^2 \right] / Q^4 \tag{9.3}$$

with Q defined by (9.2) is a conformal invariant if the derivatives of Q in (9.3) exist. We now show that q may be expressed as a function of J_{n-1} and conformal differential invariants obtained from H by the classical method.

A straightforward calculation using (4.23) shows that

$$\begin{aligned} \frac{dQ}{ds} &= H \cdot \frac{dJ}{ds} + J^2 \frac{dH}{ds}, \\ \frac{d^2Q}{ds^2} &= H \cdot \frac{d^2J}{ds^2} + 3J \frac{dJ}{ds} \frac{dH}{dS} + J^3 \frac{d^2H}{dS^2}. \end{aligned} \tag{9.4}$$

If we substitute the values given by (9.2) and (9.4) in (9.3) and use (5.9), we find

$$q = \left[H^2 \cdot J_{n-1} - 3\left(\frac{dH}{ds}\right)^2 + 2H \frac{d^2H}{dS^2} \right] / H^4. \tag{9.5}$$

This discussion shows that the only conformal scalar that one may construct by the method implicit in Theorem 5.1 and which is not obtainable by the classical method is J_{n-1} itself. To obtain J_{n-1}, we simply set $H=1$ in (9.5). Every other q is a function of J_{n-1} and conformal invariants obtained from H by the classical method.

We return to the problem of finding all conformal differential invariants of a curve. Since the value of any conformal scalar H is independent of the parametrization of the curve, we may replace t in (9.1) by a conformal arc length parameter S. In virtue of equations (6.3), the successive derivatives of x^i with respect to S may be written as functions of the conformal curvatures $J_1, J_2, \cdots, J_{r-1}, J_{n-1}$ and their derivatives with respect to S and of the g_{ij} and their derivatives with respect to the x^k and of x^i, ν^i, ${}_{(\alpha)}\eta^i$, μ^i, J, L. Since the principal normal μ^i is orthogonal to ν^i, it may be written as([57])

([57]) If $r<n-1$, we choose ${}_{(r+1)}\eta^i, {}_{(r+2)}\eta^i, \cdots, {}_{(n-1)}\eta^i$ as any normalized $(n-r-1)$-bein orthogonal to $\nu^i, {}_{(1)}\eta^i, {}_{(2)}\eta^i, \cdots, {}_{(r)}\eta^i$.

$\mu^i=\sum_\alpha a_\alpha\, {}_{(\alpha)}\eta^i$. Hence H is a function $f(\omega_1, \omega_2, \cdots, \omega_r)$ where the ω's are to be replaced by the variables:

$$x^i,\ \nu^i,\ {}_{(\alpha)}\eta^i;\ a_\alpha,\ J,\ L;\ g_{ij},\ \frac{\partial g_{ij}}{\partial x^h},\ \frac{\partial^2 g_{ij}}{\partial x^h \partial x^k},\ \cdots;$$

$$(9.6)\qquad J_1,\ \frac{dJ_1}{dS},\ \frac{d^2J_1}{dS^2},\ \cdots;\ J_2,\ \frac{dJ_2}{dS},\ \frac{d^2J_2}{dS^2},\ \cdots;\ J_{r-1},\ \frac{dJ_{r-1}}{dS},\ \frac{d^2J_{r-1}}{dS^2},\ \cdots;$$

$$J_{n-1},\ \frac{dJ_{n-1}}{dS},\ \frac{d^2J_{n-1}}{dS^2},\ \cdots.$$

Of these variables, the conformal curvatures and their derivatives with respect to S as well as ν^i, ${}_{(\alpha)}\eta^i$ are conformal geometric objects while g_{ij} and their derivatives with respect to x^k and a_α, J, L are metric geometric objects. When the enveloping space of the curve is subjected to a conformal transformation, $\overline{H}$ is the same function $f(\omega_1, \omega_2, \cdots, \omega_r)$ with the ω's replaced by the variables $\bar{x}^i$, $\bar{\nu}^i$, ${}_{(\alpha)}\bar{\eta}^i$, $\bar{a}_\alpha$, and so on, which correspond to (9.6) under the conformal mapping and $\overline{H}=H$ at corresponding points. We now prove the following theorem of which the converse statement has already been demonstrated:

THEOREM 9.1. *The most general conformal differential invariant of a curve in an $\overline{R}_n$ $(n>2)$ is a function of the conformal curvatures and their derivatives with respect to a conformal arc length parameter. Conversely, every such function is a conformal differential invariant.*

In the proof of the theorem, we consider a conformal differential invariant H defined at any point of a curve in an $\overline{R}_n$. By means of a conformal mapping of $\overline{R}_n$ on R_n, H becomes a conformal scalar defined at any point P of a curve C in R_n. We therefore first discuss invariants of C in R_n. In R_n, the coordinates x^i may be chosen so that they belong to a rectangular cartesian coordinate system U such that $g_{ij}=\delta^i_j$, $\partial g_{ij}/\partial x^h=0$, $\partial^2 g_{ij}/\partial x^h\partial x^k=0, \cdots$ throughout a region of R_n containing P and at P, $x^i=0$, $\nu^i=\delta^i_1$, ${}_{(\alpha)}\eta^i=\delta^i_{(\alpha+1)}$. If R_n is subjected to a conformal transformation[58]

$$x^i = x^i(\bar{x}^1, \bar{x}^2, \cdots, \bar{x}^n), \qquad \bar{x}^i = \bar{x}^i(x^1, x^2, \cdots, x^n)$$

belonging to G, C is mapped on another curve $\overline{C}$ in R_n and P corresponds to $\overline{P}$. In this transformation, the $\bar{x}^i$ have been chosen so that they belong to a rectangular cartesian coordinate system $\overline{U}$ such that $\bar{g}_{ij}=\delta^i_j$, $\partial\bar{g}_{ij}/\partial\bar{x}^h=0$, $\partial^2\bar{g}_{ij}/\partial\bar{x}^h\partial\bar{x}^k=0, \cdots$ throughout a region of R_n containing $\overline{P}$ and at $\overline{P}$, $\bar{x}^i=0$, $\bar{\nu}^i=\delta^i_1$, ${}_{(\alpha)}\bar{\eta}^i=\delta^i_{(\alpha+1)}$. It is clear that in the coordinate systems U and $\overline{U}$, the g_{ij} and their derivatives as well as x^i, ν^i, ${}_{(\alpha)}\eta^i$ remain unchanged by con-

(58) In this proof, we no longer assume that points with the same coordinates correspond as was done previously.

formal transformations of G. Of course the conformal curvatures and their derivatives with respect to S also remain constant. Hence in these coordinate systems, the only possible variables of H are a_α, J, L.

As a consequence of Theorem 7.1, a conformal transformation of G exists which transforms the set a_α, J, L into any other set $\bar{a}_\alpha$, $\bar{J}$, $\bar{L}$ with $J>0$, $\bar{J}>0$. Hence a_α, J, L behave like independent variables with respect to conformal transformations of G so that if any of the a_α, J, L are effective variables, H cannot remain invariant. This contradiction shows that H is independent of a_α, J and L. Hence, in the coordinate system U, H may be written as

$$
\begin{aligned}
H = H\Bigl(J_1, \frac{dJ_1}{dS}, \frac{d^2J_1}{dS^2}, \cdots ; J_2, \frac{dJ_2}{dS}, \cdots ; \\
J_{\tau-1}, \frac{dJ_{\tau-1}}{dS}, \cdots ; J_{n-1}, \frac{dJ_{n-1}}{dS}, \cdots \Bigr).
\end{aligned} \tag{9.7}
$$

Since the values of H as well as of the J's and their derivatives with respect to S do not depend on the coordinate system, (9.7) is valid in every coordinate system. (Or otherwise: The values of the g_{ij} and their derivatives and of x^i, ν^i, ${}_{(\alpha)}\eta^i$ at P behave like independent variables with respect to coordinate transformations, subject to normality conditions for ν^i, ${}_{(\alpha)}\eta^i$ while H and the conformal curvatures and their derivatives with respect to S remain constant. Therefore H cannot involve g_{ij}, x^i, ν^i, ${}_{(\alpha)}\eta^i$ as effective variables.) This observation proves the theorem for curves in R_n and conformal transformations belonging to G. But since H and the J's and their derivatives with respect to S are unchanged by a conformal mapping of R_n on an $\bar{R}_n$, (9.7) is also valid for curves in $\bar{R}_n$. The proof of Theorem 9.1 is thus complete[59].

It is clear that a similar proof in the metric theory would show that *every metric differential invariant of a curve in an R_n is a function of the metric curvatures and their derivatives with respect to a metric arc length parameter, and conversely.* We note that Theorem 9.1 is not true in general Riemann spaces. For in a space V_n whose dimensionality n exceeds 3, the Weyl conformal curvature tensor C^h_{ijk} is different from zero[60] if V_n is not an $\bar{R}_n$. Then unit vectors ${}_{(1)}\theta^i$, ${}_{(2)}\theta^i$, ${}_{(3)}\theta^i$, ${}_{(4)}\theta^i$ any two of which are either identical or mutually orthogonal exist at a point P of V_n such that $C^h_{ijk}\,{}_{(1)}\theta_h\,{}_{(2)}\theta^i\,{}_{(3)}\theta^j\,{}_{(4)}\theta^k$ is different from zero. Now, in accordance with Theorem 6.1, curves C_1 and C_2 exist in V_n whose conformal curvatures are the same functions of the conformal arc length and which pass through P so that the θ^i's are vectors of their moving conformal $(\tau+1)$-beins and whose relative conformal curvatures J at P are

---

(59) A very short non-constructive proof for the analytic case may be based on Theorem 7.2.
(60) H. Weyl, loc. cit., p. 404, and J. A. Schouten, loc. cit., p. 80.

any two different positive constants a and b. Since C^{h}_{ijk} is a conformal tensor, it follows readily from (4.22) and (4.26) that H defined by

$$H = J^{-2} C^{h}_{ijk} \, {}_{(1)}\theta_h \, {}_{(2)}\theta^i \, {}_{(3)}\theta^j \, {}_{(4)}\theta^k$$

is a conformal differential invariant of C_1 and C_2. Since $a \neq b$, H has different values at P for C_1 and C_2. Since the values of corresponding conformal curvatures and their derivatives with respect to conformal arc length parameters are the same for C_1 and C_2, H cannot be a function of the conformal curvatures and their derivatives with respect to a conformal arc length parameter. A similar example in the metric theory of a curve in a space which is not a space of constant curvature could be constructed by means of the Riemann curvature tensor R_{hijk}.

10. **Conformal null curves.** The results of the preceding sections obviously cannot be applied to a curve C whose relative conformal curvature vanishes identically. In this case the conformal arc length between any two points of C is zero, and conversely. Consequently, we call any curve along which $J \equiv 0$, a *conformal null curve*(61). As a consequence of (4.21) with $J=0$, it follows that in any V_n of class C^4 (class C^3 if V_n is conformal to an E_n), a unique conformal null curve is determined by a set of initial values for x^i, ν^i, μ^i. Hence, in contrast to the metric case, real conformal null curves exist in V_n even though its first fundamental form is positive definite. In virtue of (4.22), the conformal image of a conformal null curve is also a conformal null curve. We now derive a number of simple properties of these curves. The first of these is the theorem:

THEOREM 10.1. *A curve C in V_n $(n>2)$ is a conformal null curve if and only only if a Riemann space $\bar{V}_n$ and a conformal transformation of V_n on $\bar{V}_n$ exist such that C is mapped by this transformation on a geodesic of $\bar{V}_n$ which is also a line of principal Ricci curvature of $\bar{V}_n$.*

In the proof, we begin with the fact that a conformal transformation of V_n on some $\bar{V}_n$ exists which maps any curve (not necessarily a conformal null curve) on a geodesic of $\bar{V}_n$. In other words, every curve is conformally geodesic in some $\bar{V}_n$. The proof of this result is simple and will appear in another paper. As a consequence of this proposition, the given conformal null curve C is conformally equivalent to a geodesic $\bar{C}$ of some $\bar{V}_n$. For $\bar{C}$, $\bar{k}_1 = \bar{k}_2 = 0$. Of course the relative conformal curvature $\bar{J}$ of $\bar{C}$ is also zero. It follows from the definition of $\bar{J}^2\bar{\eta}^i$ in the equation analogous to (4.21) that

$$(10.1) \qquad \bar{R}_{hk}\bar{\nu}^k(\bar{g}^{ih} - \bar{\nu}^i\bar{\nu}^h) = 0$$

at each point of C. If we multiply this equation by $\bar{g}_{ij}$ and sum for i,

(61) In the following discussion we do not consider the non-real conformal null curves which are solutions of $ds^2 = g_{ij}dx^i dx^j = 0$.

$$[\bar{R}_{jk} - A\bar{g}_{jk}]\bar{\nu}^k = 0 \tag{10.2}$$

where $A = \bar{R}_{jk}\bar{\nu}^j\bar{\nu}^k$. Hence $\bar{\nu}^k$ is a Ricci principal direction([42]) of $\bar{V}_n$ so that $\bar{C}$ is a geodesic line of principal Ricci curvature. Conversely, if $\bar{C}$ is a curve of this kind, $\bar{k}_1 = \bar{k}_2 = 0$ and (10.2) holds. In virtue of these equations, (10.1) is true and $\bar{J}^2\bar{\eta}^i = 0$. From this equation it follows that $\bar{J} = 0$. In accordance with (4.22), J must vanish identically on any conformal image C of $\bar{C}$. This completes the proof of the theorem.

If V_n is an E_n, (6.7) holds and $R_{hk}\nu^k(g^{ih} - \nu^i\nu^h) = 0$ for every curve in E_n. Hence, as a consequence of (4.21), $J = 0$ if and only if k_1 is constant and k_2 is zero. In this case, the curve is a geodesic circle of E_n. We state these facts in this theorem:

THEOREM 10.2. *Every conformal null curve of an E_n $(n>2)$ is a geodesic circle of E_n, and conversely.*

As a result of this theorem, a curve in an $\bar{E}_n$ is a conformal null curve if and only if it is conformal to a geodesic circle in E_n. If the E_n is an R_n, any geodesic circle may be mapped on a straight line of R_n by a conformal transformation belonging to G. This proves the theorem:

THEOREM 10.3. *The necessary and sufficient condition that a curve C in an $\bar{R}_n$ $(n>2)$ be a conformal null curve is that C be the conformal image of a straight line in R_n.*

As an immediate consequence of the theorem, we note that if C_1 and C_2 are conformal null curves of two conformally euclidean spaces ${}_{(1)}\bar{R}_n$ and ${}_{(2)}\bar{R}_n$ respectively then a conformal transformation exists so that ${}_{(1)}\bar{R}_n \leftrightarrow {}_{(2)}\bar{R}_n$ and $C_1 \leftrightarrow C_2$. In other words, *all conformal null curves in $\bar{R}_n$'s are conformally equivalent.*

11. **Curves in an Einstein space E_n.** The defining equation (4.21) for the first conformal normal and the relative conformal curvature of a curve C in V_n $(n>2)$ becomes

$$J^2\eta^i = \frac{dk_1}{ds}\,{}_{(2)}\nu^i + k_1k_2\,{}_{(3)}\nu^i \tag{11.1}$$

at a point of C whenever the additional equation

$$R_{hk}\nu^k(g^{ih} - \nu^i\nu^h) = 0 \tag{11.2}$$

is satisfied. A line of reasoning similar to that employed in connection with (10.1) and (10.2) shows that (11.2) is the necessary and sufficient condition that ν^i be a Ricci principal direction of V_n. Now the second of the Frenet equations (4.13) is equivalent to

([42]) L. P. Eisenhart, loc. cit., pp. 113–114.

$$\frac{D\mu^i}{Ds} = -k_1^2 \,{}_{(1)}\nu^i + \frac{dk_1}{ds}\,{}_{(2)}\nu^i + k_1k_2\,{}_{(3)}\nu^i$$

where $\mu^i = k_1\,{}_{(2)}\nu^i$ so that (3.22) shows that

$$\frac{\mathfrak{d}\mu^i}{\mathfrak{d}s} = \frac{dk_1}{ds}\,{}_{(2)}\nu^i + k_1k_2\,{}_{(3)}\nu^i. \tag{11.3}$$

A comparison of (11.1) and (11.3) and the preceding remarks prove the theorem:

THEOREM 11.1. *The necessary and sufficient condition that*

$$J^2\eta^i = \frac{\mathfrak{d}\mu^i}{\mathfrak{d}s} \tag{11.4}$$

be true at a point of a curve of V_n $(n>2)$ is that the tangent to the curve at the point be a Ricci principal direction of V_n.

Of course, if (11.4) is true at all points of C, then C must be a line of principal Ricci curvature of V_n. It follows that (11.4) *is true for all curves of V_n if and only if V_n is an E_n*. We suppose this to be the case in the remainder of the section.

Let C be a curve in E_n. If we find the successive conformal derivatives of (11.1) with respect to s and make use of (3.22), (4.23) the Frenet equations (4.13) and the conformal Frenet equations (4.25) we obtain a series of equations of the form

$$\begin{aligned} B_1^1\,{}_{(1)}\eta^i &= A_2^1\,{}_{(2)}\nu^i + A_3^1\,{}_{(3)}\nu^i, \\ B_1^2\,{}_{(1)}\eta^i + B_2^2\,{}_{(2)}\eta^i &= A_2^2\,{}_{(2)}\nu^i + A_3^2\,{}_{(3)}\nu^i + A_4^2\,{}_{(4)}\nu^i, \\ \cdots\cdots\cdots\cdots&\cdots\cdots\cdots\cdots\cdots, \\ B_1^\omega\,{}_{(1)}\eta^i + B_2^\omega\,{}_{(2)}\eta^i + \cdots + B_\omega^\omega\,{}_{(\omega)}\eta^i &= A_2^\omega\,{}_{(2)}\nu^i + A_3^\omega\,{}_{(3)}\nu^i + \cdots \\ &\quad + A_{\omega+2}^\omega\,{}_{(\omega+2)}\nu^i. \end{aligned} \tag{11.5}$$

The B's are functions of the relative conformal curvature, the conformal curvatures (except J_{n-1}) and their derivatives while the A's are functions of the metric curvatures and their derivatives. Thus

$$B_1^1 = J^2, \qquad B_1^2 = 2J\frac{dJ}{ds}, \qquad B_2^2 = J^3J_1,$$

and

$$A_2^1 = \frac{dk_1}{ds}, \qquad A_3^1 = k_1k_2, \qquad A_2^2 = \frac{d^2k_1}{ds^2} - k_1k_2^2,$$

$$A_3^2 = 2\frac{dk_1}{ds}k_2 + k_1\frac{dk_2}{ds}, \qquad A_4^2 = k_1k_2k_3.$$

By means of equations (11.5), we may write J and the conformal curvatures as functions of the metric curvatures and their derivatives with respect to s. Thus, from the first equation, it follows that

$$J^4 = \left(\frac{dk_1}{ds}\right)^2 + (k_1k_2)^2.$$

One readily finds, if $\omega \leqq n-1$, that

$$(11.6)\qquad \begin{aligned} B^{\omega}_{\omega} &= J^{\omega+1} \cdot J_1 J_2 \cdots J_{\omega-1}, \\ B^{\omega+1}_{\omega} &= \frac{d}{ds} B^{\omega}_{\omega} + J J_{\omega-1} B^{\omega}_{\omega-1}, \\ A^{\omega}_{\omega+2} &= k_1 k_2 \cdots k_{\omega+1}, \\ A^{\omega}_{\omega+1} &= \frac{d}{ds} A^{\omega-1}_{\omega+1} + k_\omega A^{\omega-1}_{\omega}. \end{aligned}$$

We now derive a number of results which interrelate the zeros of the metric and conformal curvatures. The first of these is the theorem([68]) which follows

THEOREM 11.2. *If* $k_{\tau+1}=0$ *and either* $k_w=0$ $(w<\tau+1)$ *or* $dk_{\tau+1}/ds=0$ $(0\leqq\tau\leqq n-2)$ *at a point of a curve in* E_n $(n>2)$ *then* $JJ_1J_2\cdots J_\tau=0$ *at this point.*

The proof follows. If $JJ_1J_2\cdots J_{\tau-1}=0$ at a point P, the theorem is proved. We consider the case where $JJ_1J_2\cdots J_{\tau-1}\neq 0$ at P. Suppose that the hypothesis of the theorem is satisfied at P. Then (11.6) with $\omega=\tau, \tau+1$ shows that $A^{\tau}_{\tau+2}=A^{\tau+1}_{\tau+2}=A^{\tau+1}_{\tau+3}=0$ at P. Now the assumption $JJ_1J_2\cdots J_{\tau-1}\neq 0$ at P makes it possible to solve the first τ equations of (11.5) for ${}_{(1)}\eta^i, {}_{(2)}\eta^i, \cdots, {}_{(\tau)}\eta^i$ as vectors in the linear vector space V determined by ${}_{(2)}\nu^i, {}_{(3)}\nu^i, \cdots, {}_{(\tau+1)}\nu^i$. Since ${}_{(1)}\eta^i, {}_{(2)}\eta^i, \cdots, {}_{(\tau)}\eta^i$ are independent vectors, they may serve as a basis for V so that ${}_{(2)}\nu^i, {}_{(3)}\nu^i, \cdots, {}_{(\tau+1)}\nu^i$ may be written as linear combinations of ${}_{(1)}\eta^i, {}_{(2)}\eta^i, \cdots, {}_{(\tau)}\eta^i$. If these solutions for the ν^i's are substituted in the right member of the $(\tau+1)$st equation of (11.5), it is possible to solve for $B^{\tau+1}_{\tau+1}\,{}_{(\tau+1)}\eta^i$ as a vector in the linear vector space V. Since all the conformal normals are independent, it follows that $B^{\tau+1}_{\tau+1}=J^{\tau+2}J_1J_2\cdots J_{\tau-1}J_\tau=0$ which completes the proof of the theorem.

We now derive a result which is in the nature of a converse of the above theorem.

([68]) If $\tau=0$, we write $J_0=J$. In this case, some of the statements in the proofs of §§11 and 12 are either immediate or vacuous. The slight amendments which this necessitates can easily be supplied by the reader.

THEOREM 11.3. *If* $J_\tau=0$ $(0\leqq\tau\leqq n-2)$ *at a point of a curve in* E_n $(n>2)$ *then* $k_1k_2\cdots k_{\tau+2}=0$ *at this point.*

We first consider the case where $JJ_1J_2\cdots J_{\tau-1}\neq 0$ at the point P. Then, as above, it is possible to solve the first τ equations of (11.5) for ${}_{(1)}\eta^i, {}_{(2)}\eta^i, \cdots, {}_{(\tau)}\eta^i$ as linear combinations of ${}_{(2)}\nu^i, {}_{(3)}\nu^i, \cdots, {}_{(\tau+1)}\nu^i, {}_{(\tau+2)}\nu^i$. Since $J_\tau=0$, $B^{\tau+1}_{\tau+1}=0$ and ${}_{(\tau+1)}\eta^i$ is absent from the $(\tau+1)$st equation of (11.5). Hence, if we substitute the above solutions for ${}_{(1)}\eta^i, {}_{(2)}\eta^i, \cdots, {}_{(\tau)}\eta^i$ in the left member of the $(\tau+1)$st equation, $A^{\tau+1}_{\tau+3}\,{}_{(\tau+3)}\nu^i$ is found to lie in the linear vector space determined by ${}_{(2)}\nu^i, {}_{(3)}\nu^i, \cdots, {}_{(\tau+2)}\nu^i$. As a consequence, $A^{\tau+1}_{\tau+3}=k_1k_2\cdots k_{\tau+2}=0$ at P.

We now proceed to the case where $JJ_1J_2\cdots J_{\tau-1}=0$ at P. In this case, an $\omega\geqq 0$ exists so that $J_\omega=0$ and $JJ_1J_2\cdots J_{\omega-1}\neq 0$ at P. The discussion of the preceding paragraph then applies so that $k_1k_2\cdots k_{\omega+2}=0$ at P. The theorem is thus proved in all cases.

As a consequence of the two preceding theorems, we note this theorem:

THEOREM 11.4. *If* $k_{\tau+1}\equiv 0$ $(0\leqq\tau\leqq n-2)$ *along a curve in* E_n $(n>2)$, *then* $J_\tau\equiv 0$. *If* $J_\tau\equiv 0$ *along a curve in* E_n, *then* $k_{\tau+2}\equiv 0$.

The proof is immediate. For if $k_{\tau+1}\equiv 0$, then Theorem 11.1 shows that $JJ_1J_2\cdots J_\tau\equiv 0$. If the J's are continuous, it follows that a J_ω $(\omega\leqq\tau)$ exists which vanishes identically along the curve. Then $J_{\omega+1}=J_{\omega+2}=\cdots=J_\tau\equiv 0$. The second statement in the theorem is demonstrated in a similar manner.

We shall not prove but simply note that a theorem analogous to Theorem 12.3 exists in the case of a curve which is contained in a hypersurface E_{n-1} with indeterminate lines of curvature of an enveloping E_n.

12. **Curves in a conformally euclidean space $\overline{R}_n$.** In the metric theory of a curve in an R_n $(n>2)$, $k_{\tau+1}\equiv 0$ $(\tau\geqq 0)$ implies that the curve lies in a $(\tau+1)$-dimensional totally geodesic subspace of R_n. (The same result holds in an S_n.) We denote such a subspace which is called a $(\tau+1)$-plane by $P_{\tau+1}$. Hence $P_{\tau+1}$ is an $R_{\tau+1}$ having zero normal curvature in an enveloping R_n. If R_n is mapped conformally on itself or any other $\overline{R}_n$, then the image of $P_{\tau+1}$ in $\overline{R}_n$ is denoted by $\overline{P}_{\tau+1}$ and is called *a conformal* $(\tau+1)$*-plane*. We assume that the class of $\overline{R}_n$ is at least C^3. Now it can be shown that a subspace having umbilical points maps into a subspace with umbilical points under any conformal transformation of the enveloping space([44]). If R_n is mapped conformally on itself, then it follows that the $\overline{P}_{\tau+1}$ of R_n must be umbilical and hence are the $(\tau+1)$-dimensional spheres $S_{\tau+1}$ of R_n. This also is a consequence of the fact that every $(\tau+1)$-sphere is equivalent to a $(\tau+1)$-plane by means of a suitable inversion in R_n. Similarly, any $(\tau+1)$-dimensional subspace $V_{\tau+1}$ of an $\overline{R}_n$ having only umbilical points must be a $\overline{P}_{\tau+1}$, and conversely. For if $\overline{R}_n$ is

([44]) For example, cf. J. A. Schouten and D. J. Struik, *Einführung in die neueren Methoden der Differentialgeometrie*, vol. 2, 1938, p. 211.

mapped conformally on R_n, $V_{\tau+1}$ is transformed into an umbilical subspace of R_n, that is, an $S_{\tau+1}$. Since $S_{\tau+1}$ is a $\overline{P}_{\tau+1}$, $V_{\tau+1}$ is one also. The converse is established by reversing this argument. Thus the $\overline{P}_{\tau+1}$ of an $\overline{R}_n$ are in one-to-one correspondence with the $(\tau+1)$-spheres of R_n. In this section, some relations between the $\overline{P}_{\tau+1}$ of an $\overline{R}_n$ and the conformal curvature J_τ are derived. In particular, we obtain a conformal analogue of the theorem stated in the first sentence of this section.

A $\overline{P}_{\tau+1}$ in R_n is simply a $(\tau+1)$-sphere and is determined by $\tau+3$ points which do not lie in the same τ-sphere. By a conformal transformation of R_n on an $\overline{R}_n$, it is readily seen that, in general, a $\overline{P}_{\tau+1}$ of $\overline{R}_n$ is determined by $\tau+3$ points of $\overline{R}_n$. The *osculating conformal* $(\tau+1)$*-plane* at a point P of a curve C in $\overline{R}_n$ is a $\overline{P}_{\tau+1}$ whose order of contact with C at P is not exceeded by any other $\overline{P}_{\tau+1}$. Since $\overline{P}_{\tau+1}$ is determined by $\tau+3$ points, the order of contact of the osculating $\overline{P}_{\tau+1}$ with C is not less than $\tau+2$. If this order of contact exceeds $\tau+2$, we say that the osculating $P_{\tau+1}$ *hyperosculates* the curve C.

The discussion of the osculating $\overline{P}_{\tau+1}$ is considerably simplified by first converting this $\overline{P}_{\tau+1}$ into a $P_{\tau+1}$ in R_n. We now proceed to this simplification. Let $\overline{P}_{\tau+1}$ be the osculating conformal $(\tau+1)$-plane at a point P of a curve C in $\overline{R}_n$. Then a conformal transformation T exists so that $\overline{R}_n \leftrightarrow R_n$, $\overline{P}_{\tau+1} \leftrightarrow P_{\tau+1}$. Suppose $C \leftrightarrow C_1$, $P \leftrightarrow P_1$. Since T is continuous, the order of contact is preserved and $P_{\tau+1}$ is a $(\tau+1)$-plane in R_n whose order of contact with C_1 at P_1 is equal to that of $\overline{P}_{\tau+1}$ with C at P. We choose rectangular cartesian coordinates x^i in R_n so that P_1 is the origin of coordinates and the moving n-bein ${}_{(1)}\nu^i, {}_{(2)}\nu^i, \cdots, {}_{(n)}\nu^i$ of C_1 takes the position

$$(12.1) \qquad {}_{(1)}\nu^i = \delta^i_1, \; {}_{(2)}\nu^i = \delta^i_2, \cdots, {}_{(n)}\nu^i = \delta^i_n$$

at P. Since the order of contact of $P_{\tau+1}$ with C_1 is at least $\tau+2$, $P_{\tau+1}$ osculates C_1 (actually $P_{\tau+1}$ hyperosculates C_1 since the "normal" order of contact of an osculating $(\tau+1)$-plane with a curve is $\tau+1$) and hence must contain ${}_{(1)}\nu^i, {}_{(2)}\nu^i, \cdots, {}_{(\tau+1)}\nu^i$ at P_1(65). As a consequence of (12.1), the equations of $P_{\tau+1}$ are

$$(12.2) \qquad \begin{aligned} &x^1 = x^1, \quad x^2 = x^2, \cdots, x^{\tau+1} = x^{\tau+1}, \\ &x^{\tau+2} = 0, \; x^{\tau+3} = 0, \; \cdots, \quad x^n = 0. \end{aligned}$$

If the equations of C are $x^i = x^i(s)$ where s is a metric arc length parameter, then these equations may be written as(66)

(65) The well known fact that an osculating $P_{\tau+1}$ contains the tangent vector and the first τ normals at the point of contact need not be assumed as is done here but may readily be demonstrated using equations (12.1), (12.3) and (12.4).

(66) The analyticity of the $x^i(s)$ is not necessary as we may replace (12.3) by a finite series using the extended theorem of the mean.

$$x^i = (x^i)_0 + \left(\frac{dx^i}{ds}\right)_0 \cdot s + \left(\frac{d^2x^i}{ds^2}\right)_0 \frac{s^2}{2!} + \cdots \tag{12.3}$$

where the subscript zero signifies that the corresponding expression is to be evaluated for $x^i = 0$. Since $dx^i/ds = {}_{(1)}\nu^i$, we find by successive differentiation with respect to s and use of (4.13) that

$$\begin{aligned}
\frac{dx^i}{ds} &= D_1^1 \, {}_{(1)}\nu^i, \\
\frac{d^2x^i}{ds^2} &= D_1^2 \, {}_{(1)}\nu^i + D_2^2 \, {}_{(2)}\nu^i, \\
\frac{d^3x^i}{ds^3} &= D_1^3 \, {}_{(1)}\nu^i + D_2^3 \, {}_{(2)}\nu^i + D_3^3 \, {}_{(3)}\nu^i, \\
&\cdots\cdots\cdots\cdots\cdots\cdots, \\
\frac{d^\omega x^i}{ds^\omega} &= D_1^\omega \, {}_{(1)}\nu^i + D_2^\omega \, {}_{(2)}\nu^i + \cdots + D_\omega^\omega \, {}_{(\omega)}\nu^i.
\end{aligned} \tag{12.4}$$

The D's are functions of the metric curvatures and their derivatives. Thus $D_1^1 = 1$, $D_1^2 = 0$, $D_2^2 = k_1$, $D_1^3 = -k_1^2$, $D_2^3 = dk_1/ds$, $D_3^3 = k_1k_2$. It is a simple consequence of (4.13), that if $0 \leqq \omega \leqq n-2$,

$$\begin{aligned}
D_{\omega+2}^{\omega+2} &= k_1k_2 \cdots k_{\omega+1}, \\
D_{\omega+1}^{\omega+2} &= \frac{d}{ds} D_{\omega+1}^{\omega+1} + k_\omega \cdot D_\omega^{\omega+1}.
\end{aligned} \tag{12.5}$$

Now by definition, C_1 has contact of order w with $P_{\tau+1}$ at P_1 if the perpendicular distance from a nearby point P_2 on C_1 to $P_{\tau+1}$ is an infinitesimal of order $w+1$ with respect to the infinitesimal arc length P_1P_2. By a comparison of (12.2) and (12.3), it follows that the order of contact of C_1 with $P_{\tau+1}$ is one less than the lowest power of s which occurs in the expansions of $x^{\tau+2}(s), x^{\tau+3}(s), \cdots, x^n(s)$ in (12.3). In virtue of (12.1) and (12.4), the equations (12.3) for $i = \tau+2, \tau+3$ are

$$\begin{aligned}
x^{\tau+2} &= D_{\tau+2}^{\tau+2} \cdot \frac{s^{\tau+2}}{(\tau+2)!} + D_{\tau+2}^{\tau+3} \cdot \frac{s^{\tau+3}}{(\tau+3)!} + \cdots, \\
x^{\tau+3} &= D_{\tau+3}^{\tau+3} \cdot \frac{s^{\tau+3}}{(\tau+3)!} + \cdots,
\end{aligned} \tag{12.6}$$

where the D's are to be evaluated at the origin and it is seen that $x^{\tau+4}(s), \cdots, x^n(s)$ involve powers of s greater than $\tau+3$. Since the order of contact between C_1 and $P_{\tau+1}$ is at least $\tau+2$,

$$D^{\tau+2}_{\tau+2} = 0 \tag{12.7}$$

is a necessary condition. If $\bar{P}_{\tau+1}$ hyperosculates C, then the order of contact between C_1 and $P_{\tau+1}$ exceeds $\tau+2$, and conversely. Hence

$$D^{\tau+2}_{\tau+2} = 0, \qquad D^{\tau+3}_{\tau+3} = 0 \tag{12.8}$$

are necessary and sufficient conditions[67] that the osculating $\bar{P}_{\tau+1}$ hyperosculate C. The conditions (12.7) and (12.8) are stated in terms of the metric curvatures of C_1. We now find an equivalent statement in terms of the conformal curvatures of C_1 at P_1 and hence indirectly, in terms of the conformal curvatures of C at P.

By comparison of (11.6) and (12.5),

$$D^{\omega+2}_{\omega+2} = A^{\omega}_{\omega+2}, \qquad D^{\omega+2}_{\omega+1} = A^{\omega}_{\omega+1}. \tag{12.9}$$

According to (11.5) and (12.1), at the origin,

$$\begin{aligned} {}_{(\omega)}\eta^1 &= 0, \\ {}_{(\omega)}\eta^{\omega+3} = {}_{(\omega)}\eta^{\omega+4} = \cdots &= {}_{(\omega)}\eta^n = 0. \end{aligned} \tag{12.10}$$

It follows from (11.5), (12.1), (12.9) and (12.10) that

$$\begin{aligned} D^{\omega+2}_{\omega+2} &= B^{\omega}_{\omega}\, {}_{(\omega)}\eta^{\omega+2}, \\ D^{\omega+3}_{\omega+2} &= B^{\omega+1}_{\omega}\, {}_{(\omega)}\eta^{\omega+2} + B^{\omega+1}_{\omega+1}\, {}_{(\omega+1)}\eta^{\omega+2}. \end{aligned} \tag{12.11}$$

As a consequence of (11.6) and (12.11), the necessary condition (12.7) is equivalent to

$$J^{\tau+1} J_1 J_2 \cdots J_{\tau-1}\, {}_{(\tau)}\eta^{\tau+2} = 0, \tag{12.12}$$

and the necessary and sufficient conditions (12.8) are equivalent to

$$\begin{aligned} B^{\tau+1}_{\tau}\, {}_{(\tau)}\eta^{\tau+2} + J^{\tau+2} J_1 J_2 \cdots J_{\tau-1} J_\tau \cdot {}_{(\tau+1)}\eta^{\tau+2} &= 0, \\ J^{\tau+2} J_1 J_2 \cdots J_{\tau-1} J_\tau \cdot {}_{(\tau+1)}\eta^{\tau+3} &= 0. \end{aligned} \tag{12.13}$$

Suppose that $J J_1 J_2 \cdots J_{\tau-1} \neq 0$ at P_1. Then (12.12) shows that ${}_{(\tau)}\eta^{\tau+2} = 0$ at P_1. Then $J_\tau = 0$ is the only solution of (12.13). For, if possible, let J_τ be different from zero at P_1. Then, from (12.13), ${}_{(\tau+1)}\eta^{\tau+2} = {}_{(\tau+1)}\eta^{\tau+3} = 0$. As a consequence of (12.10), the rank of the vectors ${}_{(1)}\eta^i, {}_{(2)}\eta^i, \cdots, {}_{(\tau+1)}\eta^i$ is less than $\tau+1$ which is impossible. Hence, if $J J_1 J_2 \cdots J_{\tau-1} \neq 0$ at P_1, (12.13) is equiva-

---

(67) The discussion leading to equations (12.6) did not use the fact that $P_{\tau+1}$ hyperosculates C_1 and hence applies to the osculating $P_{\tau+1}$ of a curve C_1 in all cases. Hence (12.7) is the necessary and sufficient condition that the osculating $P_{\tau+1}$ hyperosculate C_1. If (12.8) holds as well as (12.7), then the order of contact between C_1 and $P_{\tau+1}$ is at least $\tau+3$.

lent to $J_\tau=0$ at P_1. Since the zeros of $J, J_1, J_2, \cdots, J_\tau$ for C and C_1 coincide, this completes the proof of the theorem:

THEOREM 12.1. *A conformal $(\tau+1)$-plane $(0\leqq\tau\leqq n-2)$ which osculates a curve in an $\overline{R}_n$ $(n>2)$ at a point where $JJ_1J_2\cdots J_{\tau-1}\neq 0$ hyperosculates the curve if and only if $J_\tau=0$ at the point of contact.*

Incidentally, we have shown that if $JJ_1J_2\cdots J_{\tau-1}\neq 0$ and $J_\tau=0$ at P then C in $\overline{R}_n$ is conformal to a C_1 in R_n so that (12.7) and (12.8) hold at the corresponding point P_1 of C_1. In virtue of (12.5), the solution of (12.7) and (12.8) is either

$$k_u = 0, \qquad \frac{dk_u}{ds} = 0, \qquad u \leqq \tau + 1, \tag{12.14}$$

or

$$k_w = 0, \qquad k_u = 0, \qquad w < u \leqq \tau + 1. \tag{12.15}$$

The solution (12.14) with $u<\tau+1$ is impossible under the assumption that $JJ_1J_2\cdots J_{\tau-1}\neq 0$ at P. For if $u<\tau+1$, in accordance with Theorem 11.2, $JJ_1J_2\cdots J_{u-1}=0$ at P_1 which contradicts the hypothesis. Similarly Theorem 11.2 shows that the solution (12.15) is possible only if $u=\tau+1$. These remarks prove the following converse of Theorem 11.2 for a curve in a conformally euclidean space:

THEOREM 12.2. *If $JJ_1J_2\cdots J_{\tau-1}\neq 0$ and $J_\tau=0$ $(0\leqq\tau\leqq n-2)$ at a point of a curve C in an $\overline{R}_n$ $(n>2)$, then C is conformally equivalent to a curve in R_n at whose corresponding point $k_{\tau+1}=0$ and either $dk_{\tau+1}/ds=0$ or $k_w=0$ $(w<\tau+1)$.*

If C is in a $\overline{P}_{\tau+1}$ of $\overline{R}_n$, then this $\overline{P}_{\tau+1}$ hyperosculates C at each of its points so that, in accordance with Theorem 12.2, $JJ_1J_2\cdots J_{\tau-1}J_\tau=0$ at each point of C. A proof similar to that of Theorem 11.4 shows that $J_\tau\equiv 0$ along C. Conversely, if $J_\tau\equiv 0$ along C, the osculating $\overline{P}_{\tau+1}$ of C hyperosculates C at the point of contact. However this fact need not imply that C is contained in a $\overline{P}_{\tau+1}$ since the osculating conformal $(\tau+1)$-plane may differ from point to point. We shall show that this conjecture is never actually realized and that $J_\tau\equiv 0$ implies that C lies in a $\overline{P}_{\tau+1}$ of $\overline{R}_n$.

We consider the conformal geometry of a curve C which is in a $P_{\tau+1}$ of R_n. Suppose that the equations of the $P_{\tau+1}$ in R_n which contains the curve C are

$$x^\alpha = y^\alpha, \qquad \alpha = 1, 2, \cdots, \tau + 1, \qquad x^{\tau+2} = 0, \cdots, x^n = 0,$$

where the x^i are rectangular cartesian coordinates for R_n and the y^α are the rectangular cartesian coordinates of a point of $P_{\tau+1}$. Let ϕ^i be the components in the x^i of any vector in the tangent vector space of $P_{\tau+1}$ and ψ^α the corresponding components in the y^α. Then

$$\phi^i = \psi^\alpha \cdot x^i_{,\alpha} \tag{12.16}$$

where the comma denotes covariant differentiation with respect to the y^α and the first fundamental form of P_{r+1}. In particular, the components ν^i in the x's and θ^α in the y's of the unit tangent of C are connected by the equation $\nu^i = \theta^\alpha \cdot x^i_{,\alpha}$.

Let $D\phi^i/Ds$ and $D\psi^\alpha/Ds^*$ denote the absolute derivatives of ϕ^i and ψ^α with respect to the arc length of C and the first fundamental forms of R_n and P_{r+1} respectively. If we find the absolute derivative of both members of (12.16), we obtain

$$\frac{D\phi^i}{Ds} = \frac{D\psi^\alpha}{Ds^*} \cdot x^i_{,\alpha} + x^i_{,\alpha\beta}\psi^\alpha\theta^\beta \tag{12.17}$$

where $x^i_{;\alpha\beta} = \partial^2 x^i/\partial y^\alpha \partial y^\beta$. It is clear from the defining equations of P_{r+1} that the tensor $x^i_{;\alpha\beta}$ satisfies the equation $x^i_{;\alpha\beta} = 0$. Therefore (12.17) becomes

$$\frac{D\phi^i}{Ds} = \frac{D\psi^\alpha}{Ds^*} \cdot x^i_{,\alpha}. \tag{12.18}$$

If we equate the projections in the normal vector space of C of each member of (12.18) and make use of (3.22), we obtain

$$\frac{\mathfrak{d}\phi^i}{\mathfrak{d}s} = \frac{\mathfrak{d}\psi^\alpha}{\mathfrak{d}s^*} \cdot x^i_{,\alpha}. \tag{12.19}$$

Let $\phi^i = \nu^i$, $\psi^\alpha = \theta^\alpha$ in (12.18). Then

$$\mu^i = \lambda^\alpha \cdot x^i_{,\alpha}$$

where μ^i and λ^α are the principal normals of C_1 in R_n and P_{r+1} respectively. As a result of this last equation, we may apply (12.19) with $\phi^i = \mu^i$, $\psi^\alpha = \lambda^\alpha$. If $r > 1$, after account is taken of Theorem 11.1, we obtain

$$J^2 \eta^i = J^{*2} \eta^{*\alpha} \cdot x^i_{,\alpha} \tag{12.20}$$

where J^*, $\eta^{*\alpha}$ are geometric objects of P_{r+1} analogous to J, η^i in R_n. A similar notation is used for other conformal geometric objects in P_{r+1}. We refer to geometric objects of P_{r+1} and R_n as *surface* and *space* geometric objects respectively. (We do not pause to prove that it is unnecessary to assume the existence of both the set of surface conformal objects and the set of space conformal objects. The existence of either set implies the existence of the other.)

From (12.20)

$$J = J^*, \tag{12.21}$$

$$\eta^i = \eta^{*\alpha} \cdot x^i_{,\alpha}. \tag{12.22}$$

Since the metric arc length parameters s, s^* may be chosen along C so that $s=s^*$, equations (4.23) and (12.21) lead to the conclusion that conformal arc length parameters may be chosen so that

$$S = S^*. \tag{12.23}$$

Then, from (12.19)

$$\frac{\delta\phi^i}{\delta S} = \frac{\delta\psi^*}{\delta S^*} \cdot x^i_{,\alpha}. \tag{12.24}$$

As an immediate consequence of the definition (5.10), and (12.21) and (12.23), the last surface conformal curvature J_r^* and the last space conformal curvature J_{n-1} are equal.

In virtue of (12.22), we may let $\phi^i = \eta^i$, $\psi^\alpha = \eta^{*\alpha}$ in (12.24). If we make use of the conformal Frenet equations, the resulting equation is

$$J_1\, {}_{(2)}\eta^i = J_1^*\, {}_{(2)}\eta^{*\alpha} \cdot x^i_{,\alpha} \tag{12.25}$$

from which

$$J_1 = J_1^*, \qquad {}_{(2)}\eta^i = {}_{(2)}\eta^{*\alpha} \cdot x^i_{,\alpha}. \tag{12.26}$$

We now let $\phi^i = {}_{(2)}\eta^i$, $\psi^\alpha = {}_{(2)}\eta^{*\alpha}$ in (12.24) and simplify the resulting equation by means of (4.25) and (12.25). This gives

$$J_2 = J_2^*, \qquad {}_{(3)}\eta^i = {}_{(3)}\eta^{*\alpha} \cdot x^i_{,\alpha}.$$

Proceeding in this manner, we find that the successive surface conformal normals and conformal curvatures and the corresponding space conformal normals and conformal curvatures are equal. Since only $\tau-1$ of the conformal curvatures which occur in the conformal Frenet equations of C as a curve in $P_{\tau+1}$ can be different from zero, it follows that the space conformal curvature J_τ vanishes identically.

It remains to consider the cases $\tau=0$, 1 which were excluded in the above discussion. If $\tau=0$, C is a geodesic and $J=0$. If $\tau=1$, we write $\lambda^\alpha = k_1^*\xi^\alpha$ where k_1^* is the surface first curvature and ξ^α is the surface first (metric) normal of C. If we proceed in the same way as above, we obtain

$$J^2\eta^i = \frac{dk_1^*}{ds^*}\xi^\alpha \cdot x^i_{,\alpha}, \tag{12.27}$$

$$J_1 = 0$$

instead of (12.20) and (12.25).

Conformal geometric properties similar to those derived above are enjoyed by any curve C contained in a $\bar{P}_{\tau+1}$ of an $\bar{R}_n$. For a suitable conformal transformation exists so that $\bar{R}_n \leftrightarrow R_n$, $\bar{P}_{\tau+1} \leftrightarrow P_{\tau+1}$, $C \leftrightarrow C_1$. Accordingly, the above discussion applies to C_1 in a $P_{\tau+1}$ of R_n. However, since the vectors and scalars mentioned in this discussion are the conformal normals, conformal curvatures and conformal arc length, in virtue of Theorem 4.1 and Theorem 5.2, the results apply equally well to the curve C in $\bar{P}_{\tau+1}$ in $\bar{R}_n$. These remarks complete the proof of the theorem.

THEOREM 12.3. *If a curve C in an $\bar{R}_n$ $(n>2)$ is contained in a conformal $(\tau+1)$-plane $\bar{P}_{\tau+1}$ $(0 \leqq \tau \leqq n-2)$ then the τth space conformal curvature J_τ vanishes along C. If $\tau>1$, the conformal arc length, conformal normals and conformal curvatures of C as a curve in the space $\bar{R}_n$ and as a curve in the surface $\bar{P}_{\tau+1}$ are related by the equations*:

$$
\begin{aligned}
S &= S^*, & & \\
{}_{(u)}\eta^{i} &= {}_{(u)}\eta^{*\alpha}\, x^{i}_{,\alpha}, & & u = 1, 2, \cdots, \tau, \\
J &= J^* & & \\
J_w &= J^*_w, & & w = 1, 2, \cdots, \tau - 1, \\
J_{n-1} &= J^*_\tau. & &
\end{aligned}
$$

It is now a simple matter to prove the converse theorem.

THEOREM 12.4. *If the τth conformal curvature J_τ $(0 \leqq \tau \leqq n-2)$ of a curve C in an $\bar{R}_n$ $(n>2)$ vanishes identically, then C lies in a conformal $(\tau+1)$-plane of $\bar{R}_n$.*

The proof follows. We first consider the case $\tau>1$. Let

$$J_1(S),\ J_2(S),\ \cdots,\ J_{\tau-1}(S),\ J_{n-1}(S), \tag{12.28}$$

be the conformal curvatures of C in $\bar{R}_n$. According to Theorem 6.1, a curve C_1 exists in $R_{\tau+1}$ whose conformal curvatures are the functions (12.28). If the $R_{\tau+1}$ is imbedded as a $P_{\tau+1}$ of R_n, C_1 considered as a curve in R_n will have the same functions (12.28) for its conformal curvatures as a result of Theorem 12.3. Then the conformal equivalence theorem shows that a conformal mapping exists so that $\bar{R}_n \leftrightarrow R_n$, $C \leftrightarrow C_1$. It follows that C lies in the conformal image of $P_{\tau+1}$ which is a $\bar{P}_{\tau+1}$ of $\bar{R}_n$. This completes the proof for this case.

We now consider the remaining cases, $\tau=0$ and $\tau=1$. If $\tau=0$, C is a conformal null curve and, by Theorem 10.3, C is conformal to a straight line in R_n. Hence C lies in a conformal 1-plane.

If $\tau=1$, the only conformal curvature which is not zero is $J_{n-1}(S)$. The results of §14 make it possible to give a proof for this case in precisely the same manner as was done for $\tau>1$. However, we now indicate an alternative

proof which is independent of §14. Known existence theorems for differential equations permit us to establish the existence of a solution $k_1^*(s^*)$ of the system of equations

$$2q\frac{d^2q}{ds^{*2}} - 3\left(\frac{dq}{ds^*}\right)^2 - k_1^{*2}q^2 = q^4\cdot J_{n-1}(Q), \tag{12.29}$$

$$q^2 = \left|\frac{dk_1^*}{ds^*}\right|, \tag{12.30}$$

$$Q = \int qds^*. \tag{12.31}$$

Then a curve C_1 exists in R_2 whose (metric) first curvature and (metric) arc length are k_1^* and s^* respectively. Let R_2 be imbedded as a 2-plane in R_n. According to Theorem 12.3, the only nonzero conformal curvature of C_1 as a curve in R_n is the last conformal curvature. The relative conformal curvature J of C_1 is related to k_1^* by (12.27). Then (12.30) leads to the conclusion that $J=q$. A comparison of (12.31) with (4.23) shows that $Q=S$. Now the space first curvature k_1 equals the surface curvature k_1^*. It follows from (5.9) and (12.29) that the last conformal curvature of C_1 in R_n is $J_{n-1}(S)$. According to Theorem 7.2, a conformal transformation exists so that $\overline{R}_n \leftrightarrow R_n$ $C \leftrightarrow C_1$. Hence C must lie in the conformal image of the 2-plane which contains C_1. This completes the proof of Theorem 12.4.

As an immediate consequence of Theorem 12.4, if $J_\tau=0$, C is conformal to a curve C_1 which is contained in a $(\tau+1)$-plane of R_n. Hence the $(\tau+1)$st metric curvature of C_1 vanishes. We state this result, which is similar to that of Theorem 11.4, in the theorem:

THEOREM 12.5. *If $J_\tau \equiv 0$ $(0 \leqq \tau \leqq n-2)$ along a curve C in an $\overline{R}_n$ $(n>2)$, then C is conformally equivalent to a curve in R_n whose $(\tau+1)$st metric curvature $k_{\tau+1}$ is identically zero.*

Since $\overline{P}_{\tau+1}$ in R_n is a $(\tau+1)$-sphere, we have the following corollary of Theorem 12.3 and Theorem 12.4: *The necessary and sufficient condition that a curve of R_n $(n>2)$ lie in a $(\tau+1)$-sphere of R_n is that J_τ be identically zero.* For the case $n=3$, $\tau=1$, the condition becomes $J_1=0$. This condition, when J_1 is evaluated in terms of the metric curvatures of the curve, is known[68], but the classic derivation differs completely from that of the present paper.

13. **Circular conformal transformations.** The characteristic property of the inversive group G defined in R_n is that every circle (including straight lines) of R_n is mapped on a circle under any transformation belonging to G. A generalization to any V_n of the circle in R_n is the geodesic circle defined as the curve whose first (metric) curvature is constant and whose second curva-

(68) L. P. Eisenhart, *Differential Geometry*, 1909, p. 36.

ture vanishes identically. It is therefore defined by the equations

$$\frac{dk_1}{ds} = 0, \qquad k_2 = 0. \tag{13.1}$$

The ordinary (metric) existence theorem for curves proves that a unique geodesic circle is determined by the values of ${}_{(1)}\nu^i$, ${}_{(2)}\nu^i$, k_1 at an arbitrary point of V_n. A conformal transformation of V_n on $\bar{V}_n$ which maps the geodesic circles of V_n and $\bar{V}_n$ on each other is called a *circular conformal transformation*. It is a natural generalization of the inversive transformations belonging to G. In a manner similar to that of §2, we may define tensors which have an invariant character with respect to circular conformal transformations. Analogous to the definitions in §2 we may thus define relative circular conformal tensor, circular conformetric tensor and circular conformal tensor.

According to the definition of circular conformal transformations, if (13.1) holds for a curve in V_n then $d\bar{k}_1/d\bar{s}=0$, $\bar{k}_2=0$ is true for the conformal image in $\bar{V}_n$. If these results are substituted in (4.18), we find that the mapping functions σ of the circular conformal transformations coincide with the solutions of the differential equations

$$\sigma_{hk}\nu^k(g^{ih} - \nu^i\nu^h) = 0, \tag{13.2}$$

where ν^i is an arbitrary unit vector of V_n. Equation (13.2) is analogous to (10.1). A line of reasoning similar to that employed in the discussion following (10.1) shows that ν^i in (13.2) is a principal direction determined by σ_{ij}. Since ν^i is arbitrary, it follows that *the necessary and sufficient condition that σ be the mapping function of a circular conformal transformation of V_n is that it satisfy the equations*

$$\sigma_{ij} = \phi g_{ij}. \tag{13.3}$$

Thus a V_n admits circular conformal transformations if and only if (13.3) has solutions. In a previous paper([69]), we have shown that a very large class of V_n's actually exist which admit such transformations. In particular, as follows from (4.20), (6.6) and (13.3), *any conformal transformation between Einstein spaces of dimensionality $n>2$ is circular. Conversely, the conformal image of an Einstein space under a circular conformal transformation is also an Einstein space.* A detailed study of the existence questions concerning conformal transformations between E_n's has been made by Brinkmann([70]). We also note, as was shown at the beginning of §7, *any conformal transformation between spaces of constant curvature of dimensionality $n>2$ is circular*. The converse of this statement is true even if $n=2$, that is, *if a circular conformal*

([69]) A. Fialkow, *Conformal geodesics*, loc. cit., §12.

([70]) H. W. Brinkmann, *Einstein spaces which are mapped conformally on each other*, Mathematische Annalen, vol. 94 (1925), pp. 119–145.

transformation is applied to a space of constant curvature S_n, the conformal image of S_n is also a space of constant curvature S'_n. A proof of this statement has already been given in §7 if $n>2$. The following proof also applies to the case $n=2$.

Since $S_n \leftrightarrow R_n$ by a circular conformal transformation, the conformal image spaces of S_n and R_n under all circular conformal transformations coincide. It therefore suffices to prove the above italicized statement for an R_n only. Let the x^i be rectangular cartesian coordinates. Then, if ψ is defined by (5.2), the first fundamental form of any V_n equivalent to the R_n by a circular conformal transformation is

$$ds^2 = \psi^{-2}[dx^{1^2} + dx^{2^2} + \cdots + dx^{n^2}] \tag{13.4}$$

where, as a consequence of (13.3), ψ is a solution of

$$\frac{\partial^2 \psi}{\partial x^i \partial x^j} = - \phi\psi\delta^i_j.$$

The solution of these equations exists only if ϕ may be written in the form

$$\phi = - 2a/\sum_{i=1}^{n} (ax^{i^2} + b^i x^i + c^i)$$

where a, b^i, c^i are real constants and the solution ψ is given by

$$\psi = \sum_{i=1}^{n} (ax^{i^2} + b^i x^i + c^i). \tag{13.5}$$

A comparison of (13.4) and (13.5) shows that ds^2 is the first fundamental form([71]) of an S_n whose Riemannian curvature is $\sum_{i=1}^{n}(4ac^i - b^{i^2})$. The italicized statement is thus proved.

Analogous to the present conformal theory of curves, we may develop a theory based upon circular conformal transformations. In this theory, we would restrict the conformal mapping functions to solutions of (13.3) and would consider as the enveloping space of the curve only those Riemann spaces which admit circular conformal transformations. All the results of the present theory would also hold with reference to circular conformal transformations. However, a number of new features also present themselves some of which are now indicated.

As a consequence of (4.18) and (13.3), the vector ξ^i defined by

$$\xi^i = \frac{dk_1}{ds}\,{}_{(2)}\nu^i + k_1 k_2\,{}_{(3)}\nu^i \tag{13.6}$$

is a relative circular conformal vector defined along the curve C which has the

([71]) We are here using a result which is stated in L. P. Eisenhart, *Riemannian Geometry*, 1926, p. 85.

transformation law $\bar{\xi}^i = e^{-3\sigma}\xi^i$. This vector is defined even if $n=2$. In a manner analogous to that indicated by (4.21), ξ^i determines a relative circular conformal scalar and a unit circular conformetric vector. The conformal Frenet equations for this vector and the circular conformal arc length defined with the aid of the above relative scalar yields a new sequence of "circular conformal normals" and "circular conformal curvatures." If $n>2$, the vector λ^i defined by

$$\lambda^i = J^2\eta^i - \xi^i = \frac{1}{n-2} R_{hk}\nu^k(g^{ih} - \nu^i\nu^h)$$

is also a relative circular conformal vector with the transformation law $\bar{\lambda}^i = e^{-3\sigma}\lambda^i$. Also, as follows from (4.20) and (13.3), if[72] $n>2$ any two orthogonal vectors ω^i and θ^i which are relative circular conformal vectors determine a relative circular conformal scalar Ω defined by

$$\Omega = R_{hk}\omega^h\theta^k.$$

Of course, in an E_n, $J^2\eta^i = \xi^i$ and $\lambda^i = \Omega = 0$. Similar relative circular conformal scalars and tensors could be defined using (4.19) and (13.3).

14. **The two-dimensional case.** The principal interest of circular conformal transformations lies in the possibility of utilizing them in order to develop a conformal geometry of curves in a two-dimensional Riemann space. This possibility is realized in the present section and, except where restrictive assumptions are explicitly stated, the results apply to any V_2 which admits a circular conformal transformation. Obviously the S_2's (including R_2) are such surfaces. According to an italicized statement of the preceding section, circular conformal transformations applied to surfaces of constant curvature map them on surfaces of constant curvature. Hence the theorems of this section include the circular conformal geometry of curves in surfaces of constant curvature under any circular conformal correspondence between these surfaces. In this case, as will be seen, a complete theory including the equivalence theorem is obtained. This circular conformal geometry includes the inversive theory of curves in R_2 as a special case since the transformations considered in the inversive theory are the circular conformal transformations of R_2 on itself.

But there are many V_2's besides the obvious S_2's which admit circular conformal transformations. For it may be shown, using the results of a previous paper[73], that every V_2 applicable to a surface of revolution and only these V_2's admit circular conformal transformations. The conformal image space of any of these V_2's under a circular conformal map is a surface which is also applicable to a surface of revolution. These remarks indicate the existence of a fairly large class of V_2's other than S_2's to which the present discussion applies.

---

(72) If $n=2$, $\Omega=0$.

(73) A. Fialkow, loc. cit., p. 471, equations (12.7) to (12.10) inclusive.

Any vector λ^i defined along a curve C in one of these V_2's may be resolved into tangential and normal components and written as

$$\lambda^i = \alpha\ {}_{(1)}\nu^i + \beta\ {}_{(2)}\nu^i.$$

As a result of (3.21) and the Frenet equations (4.13) for C, the conformal derivative of λ^i is defined by

$$\frac{\delta\lambda^i}{\delta t} = \frac{d\alpha}{dt}\ {}_{(1)}\nu^i + \frac{d\beta}{dt}\ {}_{(2)}\nu^i. \tag{14.1}$$

The circular conformal vector ξ^i in V_2 defined by (13.6) becomes

$$J^2\eta^i = \frac{dk}{ds}\ {}_{(2)}\nu^i$$

where, since no ambiguity is involved, we have written k for the geodesic curvature of C in V_2 instead of the usual k_1 and where similarly to (4.21), we have written $\xi^i = J^2\eta^i$. Hence the *unit circular conformetric normal* is given by

$$\eta^i = \pm\ {}_{(2)}\nu^i \tag{14.2}$$

where the algebraic sign agrees with the sign of dk/ds and the *relative circular conformal curvature* J is defined by

$$J = +\left|\frac{dk}{ds}\right|^{1/2}. \tag{14.3}$$

It follows that J is identically zero if and only if C is a geodesic circle of V_2. We exclude these curves from the present discussion. The quantity J has the transformation law (4.22) and the normal η^i or ${}_{(2)}\nu^i$ transforms according to

$${}_{(2)}\bar{\nu}^i = e^{-\sigma}\ {}_{(2)}\nu^i. \tag{14.4}$$

Indeed, since the vector space normal to ${}_{(1)}\nu^i$ is one-dimensional and is a conformal geometric object, ${}_{(2)}\nu^i$ must be a conformetric vector whose direction remains unchanged under all conformal transformations (even including the non-circular ones). As in the conformal theory, the *circular conformal arc length parameters* S are defined by (4.23) with J determined by (14.3). As a consequence of (4.23) and (14.3), $dS^2 = \pm dk\ ds$ and $JdS = \pm dk$. In virtue of (14.1) and (14.2), the circular conformal Frenet equations become the trivial equation $\delta\eta^i/\delta S = 0$ so that no "curvatures" arise in connection with the Frenet process.

We now define the circular conformal invariant of the curve C in V_2 which is analogous to J_{n-1} in the conformal theory. If we multiply (4.19) by ${}_{(1)}\nu^h\ {}_{(2)}\nu^i\ {}_{(1)}\nu^j\ {}_{(2)}\nu^k$ and sum using (3.5), (13.3), (14.4) and $g_{ij}\ {}_{(1)}\nu^i{}_{(2)}\nu^j = 0$, we obtain

$$e^{2\sigma}\bar{K} = K - 2\phi - \Delta_1\sigma \tag{14.5}$$

where K and $\bar{K}$ are the Gaussian (or Riemann) curvatures of V_2 and of its circular conformal image $\bar{V}_2$. Differentiation of (5.2) shows that $\Delta_1\psi = \psi^2\Delta_1\sigma$. This fact and equations (5.2), (5.3) and (14.5) lead to the conclusion that

$$\bar{k}^2 + \bar{K} = \psi^2(k^2 + K) + 2\psi\frac{d^2\psi}{ds^2} - \left(\frac{d\psi}{ds}\right)^2.$$

If we proceed with this equation as with (5.4), an analogue of Theorem 5.1 is obtained and the *circular conformal curvature* J_1 defined by equation (5.9) with $n=2$, $\underline{K}=K$ is a circular conformal invariant. According to (14.3),

$$J_1 = \left[4\frac{dk}{ds}\frac{d^3k}{ds^3} - 5\left(\frac{d^2k}{ds^2}\right)^2 - 4(k^2+K)\left(\frac{dk}{ds}\right)^2\right]\Big/4\left|\frac{dk}{ds}\right|^3. \tag{14.6}$$

The formally simpler quantity H defined by

$$H = \left[4\frac{dk}{ds}\frac{d^3k}{ds^3} - 5\left(\frac{d^2k}{ds^2}\right)^2 - 4(k^2+K)\left(\frac{dk}{ds}\right)^2\right]\Big/\left(\frac{dk}{ds}\right)^3$$

is obviously a circular conformal invariant, being equal to $\pm 4J_1$.

The invariant H was first obtained in the special case where $K=\bar{K}=0$ by Mullins(74) as a differential invariant of a plane curve under the group of inversions of R_2 into itself. He found H in a different form than that given above using the methods of the Lie theory. This invariant was also found by Liebmann(75), Kubota(76), Morley(77), and Patterson(78) in connection with the inversion geometry of the plane. Their methods differed from that of Mullins and the present paper, depending in most cases upon the use of the Schwarzian derivative. Some of these writers referred to S and H as the "inversive length" and "inversive curvature" respectively of a plane curve. The books of Blaschke-Thomsen(79) and Takasu(80) develop the inversive geometry of plane and space curves based upon the use of tetracyclic and pentaspherical coordinates. Recently, Maeda(81) obtained a number of new

(74) G. W. Mullins, *Differential Invariants under the Inversion Group*, Columbia University dissertation, 1917.

(75) Liebmann, *Beiträge zur Inversionsgeometrie der Kurven*, Sitzungberichte der Bayerischen Academie der Wissenschaften Munich, vol. 1 (1923), p. 79. This paper is not accessible to the author.

(76) T. Kubota, *Beiträge zur Inversionsgeometrie*, Tokyo Imperial University Science Reports, vol. 13 (1924–1925), p. 243.

(77) F. Morley, *On differential inversive geometry*, American Journal of Mathematics, vol. 48 (1926), p. 144 and *Inversive Geometry*, 1933, pp. 137–142.

(78) B. C. Patterson, *The differential invariants of inversive geometry*, American Journal of Mathematics, vol. 50 (1928), p. 556.

(79) Blaschke-Thomsen, loc. cit.

(80) T. Takasu, loc. cit.

(81) J. Maeda, *Geometric meanings of the inversion curvature of a plane curve*, Japanese Journal of Mathematics, vol. 16 (1940), pp. 177–232.

geometric interpretations of the inversive curvature of a plane curve by methods which depended upon the use of functions of a complex variable and the Schwarzian derivative.

We now enumerate those theorems of the previous sections for which analogues exist in the circular conformal geometry of V_2. A simpler phrasing of many of these theorems is possible if $n=2$. In all cases the proof in the two-dimensional case, after trivial modifications, parallels the proof already given with the understanding that $n=2$ and that S, J, J_1, η^i have the significance indicated in this section and that the word "conformal" is to be replaced by the words "circular conformal." In particular, we note that the expression $R_{jk}\nu^k(g^{ij}-\nu^i\nu^j)/(n-2)$ is to be omitted wherever it appears in a proof. After these conventions, we list the following theorems (besides those already mentioned) as having two-dimensional analogues for all surfaces applicable to a surface of revolution: Theorems 5.2, 6.1, 10.2. In Theorem 6.1, the class of V_2 may be C^3 instead of C^4. In Theorem 10.2, the condition that V_2 be an E_2 is universally true and should be omitted. The following theorems have analogues only if V_2 is an S_2: Theorems 7.1, 7.2, 8.1, 8.2, 8.3, 9.1, 10.3. The theorems of §11 are trivially true for curves in V_2 and the theorems of §12 are trivially true for curves in S_2. We note however as a consequence of (12.27), (14.3) and (5.9) with $n=2$ (or the equivalent equation (14.6)) that Theorem 12.3 has the following analogue if $\tau=1$ (using the notation of the theorem): *If a curve C in an* $\bar{R}_n$ *(or* S_n*) is contained in a 2-sphere* $\overline{P}_2$ *of* R_n *(or* S_n*) then the first space conformal curvature* J_1 *vanishes along C and the following equations hold*:

$$S = S^*, \qquad J = J^*, \qquad J_{n-1} = J_1^*$$

where S^*, J^*, J_1^* *are the circular conformal arc length, relative curvature and curvature respectively of C in* $\overline{P}_2$. This result may now be used to prove Theorem 12.4 for the case $\tau=1$.

The equivalence theorem for curves in the inversion geometry of the plane was proved in a different form by Kubota and stated in the present form by Morley and Patterson. The plane curves along which H is a constant were studied by Mullins who showed that they are the inversive images of the logarithmic spirals. According to the two-dimensional analogue of Theorem 8.3, they are the paths of the inversive group.

In this section, we have developed the circular conformal theory of curves in a V_2. While a curve has circular conformal invariants, it cannot have any conformal invariants since every analytic curve in a V_2 is conformally equivalent to a straight line in R_2. However the horn angle formed by two tangent curves C_1 and C_2 in a V_2 does have a conformal invariant which we now derive. Since $n=2$, $k_2=0$ so that, as before, we may write k for k_1. Also

$$g^{ij} = {}_{(1)}\nu^i\,{}_{(1)}\nu^j + {}_{(2)}\nu^i\,{}_{(2)}\nu^j. \tag{14.7}$$

If we use (4.13), (14.4) and (14.7), then (3.6) is equivalent to

$$\bar{k} = e^{-\sigma}(k - \sigma_{,h}\nu^h). \tag{14.8}$$

Similarly (4.18) becomes, after account is taken of (14.4) and (14.7)

$$\frac{d\bar{k}}{d\bar{s}} = e^{-2\sigma}\left(\frac{dk}{ds} - \sigma_{hk\ (2)}\nu^h{}_{(1)}\nu^k\right). \tag{14.9}$$

It follows from (14.8) and (14.9) that

$$\left[\frac{dk_{(1)}}{ds_{(1)}} - \frac{dk_{(2)}}{ds_{(2)}}\right] \Big/ [k_{(1)} - k_{(2)}]^2,$$

where the subscripts (1) and (2) refer to the geometric objects of C_1 and C_2 respectively, is a conformal invariant of the horn angle. It is the measure of the horn angle discovered by Kasner([82]) for the plane case who also proved that it is sufficient for a conformal characterization of the horn angle (except possibly for invariants of infinite order). Kasner's results have been extended to any two-dimensional Riemann space by Comenetz([83]). A very detailed geometry of horn angles based upon the above measure has been developed by Kasner([84]). We note that the method of this paper may be utilized to obtain conformal invariants of a horn angle in any V_n $(n>2)$. However the results are not as significant as in the two-dimensional case since each of the constituent curves of the horn angle has conformal properties of $n>2$.

15. **Curves in a conformal Riemann space V_n.** In this section, we show how our previous results may be used to develop a theory of curves in a V_n which is based upon the tensor $g_{ij}/g^{1/n}$. Following T. Y. Thomas, we define the conformal Riemann space V_n of class C^m as the space whose coordinate manifold is of class C^m and whose fundamental geometric object G_{ij}, defined over the manifold, is of class C^{m-1}. The tensor G_{ij} is a symmetric, positive definite relative tensor of weight $-2/n$ with respect to coordinate transformations, that is,

$$G'_{ij} = \Delta^{-2/n} G_{hk} \frac{\partial x^h}{\partial x'^i} \frac{\partial x^k}{\partial x'^j} \tag{15.1}$$

where Δ is the Jacobian $|\partial x^i/\partial x'^j|$ of the transformation. It follows from (15.1) that $G = |G_{ij}|$ is a scalar, that is,

---

([82]) E. Kasner, *Conformal geometry*, Proceedings of the Fifth International Congress of Mathematicians, vol. 2 (1912), p. 81.

([83]) G. Comenetz, *Conformal geometry on a surface*, Annals of Mathematics, (2), vol. 39 (1938), pp. 863–871.

([84]) E. Kasner, *Trihornometry: A new chapter of conformal geometry*, Proceedings of the National Academy of Sciences, vol. 23 (1937), pp. 337–341 and *Fundamental theorems of trihornometry*, Science, vol. 85 (1937), pp. 480–482.

$$G' = G. \tag{15.2}$$

For reasons stated in §2, we restrict ourselves to a portion of V_n which is a neighborhood U of a point coverable by a single coordinate system $\{x^i\}$. We shall continue to refer to U as the conformal Riemann space V_n. Let F denote the set of all positive functions Ω of class C^{m-1} defined over V_n which are relative scalars with respect to coordinate transformations having the transformation law

$$\Omega' = |\Delta|^{1/n} \cdot \Omega. \tag{15.3}$$

If Ω belongs to F, then the geometric object g_{ij} defined by

$$g_{ij} = \Omega^2 G_{ij} \tag{15.4}$$

is a tensor, that is,

$$g'_{ij} = g_{hk} \frac{\partial x^h}{\partial x'^i} \frac{\partial x^k}{\partial x'^j}.$$

We note as a consequence of (15.4) that

$$\Omega = g^{1/2n}/G^{1/2n}.$$

If $\overline{\Omega}$ is any other scalar in F then

$$\bar{g}_{ij} = \overline{\Omega}^2 G_{ij} \tag{15.5}$$

is also a tensor. Also, as follows from (15.3), e^σ defined by

$$e^\sigma = \frac{\overline{\Omega}}{\Omega} \tag{15.6}$$

is an absolute scalar with respect to coordinate transformations. As a consequence of (15.4), (15.5) and (15.6),

$$\bar{g}_{ij} = e^{2\sigma} g_{ij}. \tag{15.7}$$

Let U and $\overline{U}$ be any two coordinate neighborhoods of class C^m whose points correspond to those of U by means of point transformations which are of class C^m in the local coordinates of the neighborhoods. Then, as was shown in §2, allowable coordinate systems $\{x^i\}$ may be chosen in U and $\overline{U}$ so that corresponding points of U and U and of $\overline{U}$ and U have the same coordinates. Throughout this discussion, we assume that the coordinates $\{x^i\}$ are chosen so that points with the same coordinates correspond. In these coordinate systems, the tensors g_{ij} and $\bar{g}_{ij}$ defined over U and $\overline{U}$ respectively determine two Riemann spaces V_n and $\overline{V}_n$ of class C^m whose respective metric tensors they are. According to (15.7), the induced transformation which maps points of U

and $\bar{U}$ with the same coordinates is a conformal correspondence of the points of V_n and $\bar{V}_n$. As Ω ranges over all values of the set F, the corresponding σ's range over all functions of class C^{m-1} over U and the $\bar{g}_{ij}$ are any tensors related to one of them by (15.7). Associated with V_n is the set of conformal correspondences of class C^m whose domains are the $\bar{V}_n$. We denote this set by[36] Ψ.

Corresponding to any geometric object in V_n, there exists a set of geometric objects, one in each Riemann space $\bar{V}_n$, which are in conformal correspondence by means of the transformations of Ψ. For example, corresponding to a tensor $T^{i_1\cdots i_p}_{j_1\cdots j_r}$ in V_n are the set of conformal tensors in the $\bar{V}_n$ whose components coincide with those of $T^{i_1\cdots i_p}_{j_1\cdots j_r}$. Conversely, every conformal tensor in the $\bar{V}_n$ determines a unique tensor in V_n. Indeed relative conformal tensors in the $\bar{V}_n$ also define tensors in V_n. For, in accordance with the remarks of §2, every relative conformal tensor in $\bar{V}_n$ corresponds to a unique conformal tensor if a relative conformal scalar exists in the space. As a consequence of these remarks, any theorem concerning conformal geometric objects in the $\bar{V}_n$ which is independent of the particular mapping function $\sigma(x^i)$, and hence depends only on the set F and not on the particular function Ω belonging to F, is also a theorem about geometric objects of V_n. These observations apply to the conformal theory of curves in conformally equivalent Riemann spaces which is developed in the previous sections of the paper. Consequently the previous results also constitute a theory of curves in V_n. If the Weyl conformal curvature tensor C^h_{ijk} of V_n vanishes, V_n is a flat conformal space. In this case, the $\bar{V}_n$ are the conformally euclidean spaces $\bar{R}_n$ which are related by the correspondences of Ψ and the previous theorems lead to a complete theory of curves in V_n.

In what follows, we give the outline of the theory of curves which is based (formally) upon the tensor G_{ij}. If we write $e^{\phi}=\Omega$, then (15.4) becomes $g_{ij}=e^{2\phi}G_{ij}$ which is analogous to (2.6) with σ, g_{ij}, $\bar{g}_{ij}$ replaced by ϕ, G_{ij}, g_{ij} respectively. Of course the analogy is not complete since ϕ is not a scalar and G_{ij} is not a simple tensor with respect to coordinate transformations. However, this complication does not affect the argument which follows since we remain in the same coordinate system. However, the geometric objects which we define (arc length, curvatures, normals) are absolute scalars and vectors with respect to coordinate transformations.

Let

$$x^i = x^i(s)$$

---

[36] We may assume that G is constant. If G were a nonconstant function $A(x^i)$, the equivalence theory for V_n would be reducible to the corresponding theory for a Riemann space. For, as a consequence of (15.2) and (15.7), the tensor f_{ij} defined by

$$f_{ij} = [g^{hk}A_{,h}A_{,k}]\cdot g_{ij}$$

has the same components for every Ω belonging to F. The equivalence theory of V_n is therefore the equivalence theory of the Riemann space whose metric tensor is f_{ij}.

be the equations of a curve $\boldsymbol{C}$ in $\boldsymbol{V}_n$. We define

$$\left\{ {i \atop jk} \right\}$$

as the Christoffel symbols of the second kind formed from G_{ij} and θ^i, ξ^i by the equations

$$\theta^i = \frac{dx^i}{ds}$$

$$\xi^i = \frac{d^2x^i}{ds^2} + \left\{ {i \atop jk} \right\} \frac{dx^j}{ds} \frac{dx^k}{ds}$$

where s is determined by

$$G_{ij} \frac{dx^i}{ds} \frac{dx^j}{ds} = 1.$$

Hence θ^i and ξ^i are the formal analogues of the unit tangent and principal normal of the corresponding curve C in V_n. The "relative curvature" $\boldsymbol{J}$ and the vector ${}_{(1)}\boldsymbol{n}^i$ of $\boldsymbol{C}$ are defined by an equation analogous to (4.21) so that, as in (4.22),

$$(15.8) \qquad \boldsymbol{J} = e^{-\phi} J, \qquad {}_{(1)}\eta^i = e^{-\phi}\, {}_{(1)}\boldsymbol{n}^i$$

where the geometric objects J and ${}_{(1)}\eta^i$ refer to the curve C in V_n. Therefore $\boldsymbol{S}$, defined by

$$\boldsymbol{S} = \int \boldsymbol{J} d\boldsymbol{s}$$

is equal to the integral invariant S given by (4.23). Since S remains unchanged under transformations of coordinates, this is also true of $\boldsymbol{S}$ so that $\boldsymbol{S}$ is a scalar. It plays the role of an "arc length parameter" for the curve $\boldsymbol{C}$ in $\boldsymbol{V}_n$.

We define ${}_{(1)}\boldsymbol{\lambda}^i$ by ${}_{(1)}\boldsymbol{\lambda}^i = \boldsymbol{J}^{-1}\, {}_{(1)}\boldsymbol{n}^i$. It follows from (15.8) that ${}_{(1)}\boldsymbol{\lambda}^i = {}_{(1)}\lambda^i$ where ${}_{(1)}\lambda^i$ is defined by (4.27). Since ${}_{(1)}\lambda^i$ is a conformal vector, ${}_{(1)}\boldsymbol{\lambda}^i$ must transform like a vector under coordinate changes. It is the "first normal" of $\boldsymbol{C}$.

Let $\boldsymbol{\Gamma}^i_{jk}$ be defined by an equation similar to (3.29). Then

$$\boldsymbol{\Gamma}^i_{jk} = \left\{ {i \atop jk} \right\} + \left(\xi_j + \theta_j \frac{d \log \boldsymbol{J}}{ds} \right) \delta^i_k + \left(\xi_k + \theta_k \frac{d \log \boldsymbol{J}}{ds} \right) \delta^i_j$$
$$- G_{jk} \left(\xi^i + \theta^i \frac{d \log \boldsymbol{J}}{ds} \right).$$

Then, as in (3.31), it follows that $\boldsymbol{\Gamma}^i_{jk} = \Gamma'^i_{jk}$. Since the Γ'^i_{jk} transform like coefficients of connection under coordinate transformations, this is also true for

the Γ^i_{jk}. We use these Γ^i_{jk} to define the "derivative" of any tensor $T^{i_1\cdots i_p}_{j_1\cdots j_q}$ in a manner analogous to the definition given by (3.30). Consequently, this derivative has properties similar to those described in (A′), (B), (C), (D′) and (E′). In particular, the analogue of (A′) states that the Γ^i_{jk} derivative of any tensor $T^{i_1\cdots i_p}_{j_1\cdots j_q}$ transforms like a tensor of the same kind under coordinate transformations.

If this derivative is applied successively to ${}_{(1)}\lambda^i$ a sequence of equations analogous to (4.28) is obtained. They are

$$\frac{\delta_{(\alpha)}\lambda^i}{\delta S} = - J_{\alpha-1}\,{}_{(\alpha-1)}\lambda^i + J_\alpha\,{}_{(\alpha+1)}\lambda^i \quad J_0 = J_r = 0, \qquad \alpha = 1, 2, \cdots, r,$$

and are the Frenet equations of the curve C. The proofs of §4 show that the J_α are scalars and that the ${}_{(\alpha)}\lambda^i$ are vectors with respect to coordinate transformations in V_n. They are the "curvatures" and "normals" of C. The "$(n-1)$st curvature" J_{n-1} may be obtained as in §5 and the other results of this paper also have application here.

We note that the $\delta/\delta S$ process of differentiation defined by means of the Γ^i_{jk} is with respect to and depends upon the curve C in V_n and is therefore not appropriate for the purpose of obtaining a characterization of the entire space V_n (unless one could define a congruence of curves intrinsically in V_n). While the applicability of the derivative appears limited in this sense, its simple structure and conformal properties noted in §3 make it a suitable tool in the theory of curves.

BROOKLYN COLLEGE,
BROOKLYN, N. Y.

AN ARITHMETICAL THEORY OF THE BERNOULLI NUMBERS

BY

H. S. VANDIVER

In the present paper we shall describe a method which enables us to find many new types of relations concerning the Bernoulli and allied numbers. The scheme might be described as ultra-arithmetical in character. It depends mainly on the following idea. Let a and b be rational with $a \equiv b$, modulo p, where p is any prime integer. If a and b do not depend on p, it then follows, since there is an infinity of primes, that $a = b$.

A similar method has been employed in other parts of mathematics; for example, Hasse(1) in a paper on algebraic geometry uses the method and comments upon the success it has had in various lines.

Perhaps the simplest looking formula in which a Bernoulli number appears alone on one side of the relation is as follows, if $S_n(p) = 1^n + \cdots + (p-1)^n$,

$$\frac{S_n(p)}{p} \equiv b_n \pmod{p},$$

where $n+1 < p$, in which case, of course, the left-hand member of the congruence is an integer. In order to take advantage of this simplicity we employ extensively the function which we have called in a previous paper the Mirimanoff polynomial(2), namely, the relation (1) which follows. This is connected with the previous congruence, if we note that

$$f_n^{(p)}(1) = S_n(p).$$

In general we employ more or less obvious identities involving one or more indeterminates, then operate thereon, using the method of formal exponential differentiation explained in another paper(3). The elementary function from which the Mirimanoff polynomials are generated by this process is

$$\frac{x^m - 1}{x - 1}.$$

Presented to the Society, September 5, 1941; received by the editors April 5, 1941.

(1) Abhandlungen Göttingen, vol. 18 (1937), pp. 51–55; cf. also Vandiver, Bulletin of the American Mathematical Society, vol. 31 (1925), p. 348; in particular, the proof of II. There is a misprint in the first congruence involving H. The right-hand member should read a_e in lieu of a^e.

(2) Vandiver, Duke Mathematical Journal, vol. 3 (1937), p. 570; so-called because Mirimanoff, it appears, first investigated its properties extensively in an article in Crelle's Journal, vol. 128 (1905), pp. 45–68.

(3) Vandiver, *On formal exponential differentiation in rings*, Proceedings of the National Academy of Sciences, vol. 28 (1942), pp. 24–27.

In view of the congruence (68) below it seems to me that the theory of Mirimanoff polynomials has often been obscured by the theory of the Euler polynomials defined in (12) and (13). I think that this is unfortunate, as the Mirimanoff polynomial has a much simpler algebraic form. Each type is generalized in §2 of this paper. In (19) instead of a congruence we derive an equation involving the two types of polynomials.

If in place of the simple congruence involving $S_n(p)$ given above we employ a known congruence such as the following:

$$\frac{(n^i - 1)b_i}{i} \equiv y_a a^{i-1} \pmod{p},$$

where $y_a \equiv -a/p \pmod{n}$, we are forced to use an extension of the Mirimanoff polynomials and relations involving the number

$$\frac{(m^n - 1)b_n}{n}$$

in lieu of b_n itself. The former number has appeared in a great number of investigations concerning Bernoulli numbers, but its properties seem to be quite different in many connections; for example, I quote Frobenius as follows: " . . . die Tangentenkoefficienten deren Theorie man in den bisherigen Darstellungen nicht scharf genug von der eigentlichen Bernoullischen Zahlen geschieden hat."

In a previous paper(4) the writer introduced the idea of Bernoulli numbers of various orders such that we call $b_n(m, k) = (mb+k)^n$, $m \neq 0$, *a generalized Bernoulli number of the first order*; and a number of the form, for $r > 1$,

$$(m_r b^{(r)} + m_{r-1} b^{(r-1)} + \cdots + m_1 b' + m_0)^n = b_n(m_r, m_{r-1}, \cdots, m_0),$$

where this expression is expanded in full by the multinomial theorem and b_a substituted for $b_a^{(i)}$, $i = 1, 2, \cdots, r$, and where the m's are integers, $m_i \neq 0$, *a Bernoulli number of the rth order*. This is an extension of the definition of Lucas of the ultra-Bernoulli numbers(5).

Bernoulli numbers of the first order are considered in §§3 to 11 of this paper. In §5 a generalization of the von Staudt-Clausen theorem is derived which applies to Bernoulli numbers of the first order (Theorem I). In another

(4) Vandiver, Proceedings of the National Academy of Sciences, vol. 23 (1937), p. 555.

(5) Frobenius, *Berlin Sitzungsberichte*, 1910, p. 810. "Die Bezeichnung der Werte

$$(b + b')^n, \cdots, (b + b^{(1)} + b^{(2)} + \cdots + b^{(n)})^n,$$

als Bernoullische Zahlen höherer Ordnung oder gar als ultra-bernoullische Zahlen scheint mir wenig glücklich gewählt und mehr von abschreckender Wirkung zu sein."

The writer differs from Frobenius regarding this. These numbers, as well as the generalizations of them considered in this paper are shown to be natural analogues of the ordinary Bernoulli numbers.

paper(6) a generalization of certain congruence properties of the numerators of Bernoulli numbers is obtained which applies to the numerators of any Bernoulli number of the first order. It is noteworthy that these generalizations are very little more complicated in statement than those theorems which apply to the ordinary Bernoulli numbers.

In order to illustrate the varied applications of our method and to develop a connected theory, we give new proofs of several known theorems. In particular, although Theorem I was proved in previous papers(6), we give two new proofs of the same, as the result and the material in the proofs are both important in our theory.

The properties of the Bernoulli numbers of the second order are considered in §11 to §15. It seems to me that from many standpoints this type of Bernoulli number is the most remarkable. For example, we note from Theorem III that the only prime factors occurring in the denominator of such a number, say $(kb+jb'+h)^n$, n even, are divisors of jk and are also found among the von Staudt-Clausen primes of order n. As noted in §16, this property does not carry over to Bernoulli numbers of higher order. Also (Theorem IV, Corollary I), any Bernoulli number of the above type can be expressed as a linear combination of Bernoulli numbers of the first order with coefficients whose denominators divide the integers k and j. In particular cases the equivalent of this result may be expressed in terms of the roots of unity (Theorem V). Results of an entirely new type are given in Corollary I and Corollary II to Theorem V.

Since congruence methods are employed throughout this paper, one might imagine that many new congruences concerning the Bernoulli numbers could be obtained aside from those given here. Such is indeed the case, but their statement will be reserved for other papers.

1. **Euler and Mirimanoff polynomials.** We write the Mirimanoff polynomials in the form

$$f_a^{(m)}(x) = 0^a + 1^a x + 2^a x^2 + \cdots + (m-1)^a x^{m-1}, \tag{1}$$

where $0^0 = 1$.

We shall first show that, for $n > 0$,

$$x[(f(x) + 1)^n \equiv f_n(x) \pmod{p}, \tag{2}$$

where $f_i(x) = f_i^{(p)}(x)$, p is prime, and the left bracket symbol in the left-hand member signifies that the nth power of $(f(x)+1)$ is to be taken symbolically, that is, after development by the binomial theorem the exponents are de-

(6) Vandiver, *On simple explicit expressions for generalized Bernoulli numbers of the first order*, Duke Mathematical Journal, vol. 8 (1941), pp. 575–584; Carlitz, *Generalized Bernoulli and Euler numbers*, ibid., pp. 586–587.

graded to subscripts and, in particular, the last term is expressed as $f_0(x)$. Also for $n=0$ any such expression is taken as unity.

Consider

$$(3)\qquad x(1+x+x^2+\cdots+x^{p-1}) = x+x^2+\cdots+x^p.$$

Now employ formal exponential differentiation (Vandiver[3])). Differentiating this expression n times with respect to x, we obtain

$$(4)\qquad \begin{aligned} x[(f(x)+1)^n &= x+2^nx^2+\cdots+(p-1)^nx^{p-1}+p^nx^p \\ &= f_n(x)+p^nx^p, \end{aligned}$$

from which (2) follows.

Now consider

$$(5)\qquad x+x^2+\cdots+x^p = (1-x^p)x/(1-x).$$

Differentiate this n times, where $\bar{d}$ denotes exponential differentiation; we find

$$x+2^nx^2+\cdots+p^nx^p = (1-x^p)\left[\frac{\bar{d}^n}{dx^n}\left(\frac{x}{1-x}\right)\right]+\frac{pF(x)}{(1-x)^n},$$

$$f_n(x)+p^nx^p = (1-x^p)\left[\frac{\bar{d}^n}{dx^n}\left(\frac{x}{1-x}\right)\right]+\frac{pF(x)}{(1-x)^n},$$

or

$$(6)\qquad f_n(x) = (1-x^p)\left[\frac{\bar{d}^n}{dx^n}\left(\frac{x}{1-x}\right)\right]+p\,\frac{G(x)}{(1-x)^n}.$$

We then have, for $n>0$,

$$(7)\qquad f_n(x) \equiv (1-x^p)\left[\frac{\bar{d}^n}{dx^n}\left(\frac{x}{1-x}\right)\right] \pmod{p},$$

where we write $M(x)/N(x)\equiv 0 \pmod{p}$, if $M(x)$ and $N(x)$ are polynomials in which each coefficient of $M(x)$ is divisible by p, while not all the coefficients of $N(x)$ are so divisible.

We write

$$H_n(x) = \frac{1-x}{x}\left[\frac{\bar{d}^n}{dx^n}\left(\frac{x}{1-x}\right)\right],$$

so that

$$(8)\qquad f_n(x) \equiv (-1)^n\frac{(1-x^p)x}{1-x}H_n(x) \pmod{p}.$$

Now consider

$$x^{-1}(1+x+x^2+\cdots+x^{p-1}) = 1+x+x^2+\cdots+x^{p-1}-x^{p-1}+x^{-p}\cdot x^{p-1}.$$

Differentiating n times, we obtain

$$\frac{1}{x}\,[(f(x) - 1)^n \equiv f_n(x) + (p-1)^n x^{p-1} x^{-p} - (p-1)^n x^{p-1} \pmod{p},$$

whence

$$(9) \qquad [(f(x) - 1)^n \equiv x f_n(x) + (-1)^n (1 - x^p) \pmod{p}.$$

We now express each $f_a(x)$ in this relation in terms of $H_a(x)$ by using (8), and this gives(7), modulo p,

$$(10) \qquad x\,\frac{1-x^p}{1-x}\left[(-H-1)^n - (-1)^n + (-1)^n \frac{1-x^p}{1-x}\right. \equiv (-1)^n x^2 \frac{1-x^p}{1-x} H_n + (-1)^n (1-x^p),$$

or

$$(11) \qquad (-1)^n\left[x\,\frac{1-x^p}{1-x}\left\{[(H+1)^n - 1\right\} + \frac{1-x^p}{1-x}\right] \equiv (-1)^n\left[x^2 \frac{1-x^p}{1-x} H_n + (1-x^p)\right].$$

Divide (11) by $(-1)^n$ and $(1-x^p)$ and multiply by $1-x$ to get

$$x[(H+1)^n - x + 1 \equiv x^2 H_n + 1 - x \pmod{p},$$

or

$$x[(H+1)^n \equiv x^2 H_n \pmod{p},$$

whence

$$[(H+1)^n \equiv x H_n \pmod{p}.$$

But this relation is independent of p, an arbitrary prime, hence is an equation, and we have(8)

$$(12) \qquad [(H+1)^n = x H_n, \qquad n > 0.$$

Taking $H_0 = 1$, we obtain,

$$H_1 = 1/(x-1),$$
$$H_2 = (1+x)/(x-1)^2,$$

(7) In (10) we substitute the H_0 terms $(-1)^n$ of the $(-H-1)^n$ expansion and add $f_0 = (1-x^p)/(1-x)(-1)^n$ of the $(f+1)^n$ expansion, because (8) does not hold for $n=0$.

(8) Frobenius, *Berlin Sitzungsberichte*, 1910, p. 828.

$$H_3 = (1 + 4x + x^2)/(x - 1)^3,$$
$$H_4 = (1 + 11x + 11x^2 + x^3)/(x - 1)^4,$$
$$H_5 = (1 + 26x + 66x^2 + 26x^3 + x^4)/(x - 1)^5.$$

Now write

$$H_n = R_n(x)/(x - 1)^n. \tag{13}$$

The $R_n(x)$ are the *Euler polynomials*,

$$R_1 = 1,$$
$$R_2 = 1 + x,$$
$$\cdots\cdots$$

Putting (8) in (2) we obtain

$$x^2 \frac{1 - x^p}{1 - x} \{ [(- H + 1)^n - 1 \} + x \frac{1 - x^p}{1 - x} \equiv (- 1)^n x \frac{1 - x^p}{1 - x} H_n \pmod{p},$$
$$x^2 \{ [(- H + 1)^n - 1 \} + x \equiv (- 1)^n x H_n \pmod{p},$$
$$x^2 \{ [(H - 1)^n - (- 1)^n \} + (- 1)^n x \equiv x H_n, \tag{14}$$
$$x [(H - 1)^n - x(- 1)^n + (- 1)^n \equiv H_n \pmod{p},$$
$$x [(H - 1)^n - x(- 1)^n + (- 1)^n = H_n,$$

which is another recursion formula for the H's.

We also have

$$\begin{aligned} [(k + (H + 1)^n) &= k^n + C_{n,1} k^{n-1}(H + 1) \\ &\quad + C_{n,2} k^{n-2}(H + 1)^2 + \cdots + (H + 1)^n \\ &= k^n + C_{n,1} k^{n-1} x H_1 + C_{n,2} k^{n-2} x H_2 + \cdots + x H_n, \end{aligned} \tag{15}$$

and

$$[(k + H)^n = k^n + C_{n,1} k^{n-1} H_1 + C_{n,2} k^{n-2} H_2 + \cdots + H_n. \tag{16}$$

Multiply (16) by x and subtract from (15). We then obtain(8)

$$[(k + H + 1)^n - x[(k + H)^n = k^n(1 - x). \tag{17}$$

Setting $k = 0, 1, 2, \cdots, r$ in (9), we obtain

$$[(H + 1)^n - x H_n = 0,$$
$$[(H + 2)^n - x[(H + 1)^n = 1^n(1 - x),$$
$$[(H + 3)^n - x[(H + 2)^n = 2^n(1 - x),$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,$$
$$[(H + r)^n - x[(H + r - 1)^n = (r - 1)^n(1 - x).$$

Now multiply the first of these by x^{r-1}, the second by x^{r-2}, and so on, and add. We then obtain

$$
\begin{aligned}
[(H+r)^n - x^r H_n &= (1-x)\big[1^n x^{r-2} + 2^n x^{r-3} + 3^n x^{r-4} + \cdots \\
&\qquad + (r-2)^n x + (r-1)^n\big] \\
&= (1-x) x^{r-1} f_n^{r}\left(\frac{1}{x}\right).
\end{aligned} \tag{18}
$$

Using a similar scheme, we obtain[9] from (14)

$$
x^{r+1}[(H-r)^n - xH_n = (-1)^n (x-1) f_n^{(r+1)}(x). \tag{19}
$$

This is a rather curious relation; the expression on the left involves quotients of polynomials, while $f_n(x)$ is a polynomial for each r.

2. **Generalization of the Mirimanoff and Euler polynomials and a related formula.** Set

$$
\begin{aligned}
f_a(x, m, k) = k^a &+ (k+m)^a x + (k+2m)^a x^2 + \cdots \\
&+ (k + (p-1)m)^a x^{p-1}.
\end{aligned} \tag{20}
$$

We now consider

$$
\frac{y^k(1 - y^{mp}x^p)}{1 - y^m x} = y^k + y^{(k+m)} x + \cdots + y^{(k+(p-1)m)} x^{p-1}. \tag{21}
$$

Differentiate this expression a times exponentially with respect to y and set $y=1$. We then obtain (20) on the right; on the left we differentiate in the form

$$
\frac{y^k(1 - y^{mp}x^p)}{1 - y^m x} = (1 - y^{mp}x^p)\left(\frac{y^k}{1 - y^m x}\right), \tag{22}
$$

and we then obtain

$$
f_n(x, m, k) \equiv (1 - x^p)\left[\frac{\overline{d^n}}{dy^n}\left(\frac{y^k}{1 - xy^m}\right)\right]_{y=1} \pmod{p}. \tag{23}
$$

We now set

$$
\left[\frac{\overline{d^n}}{dy^n}\left(\frac{y^k}{1 - xy^m}\right)\right]_{y=1} = \frac{\overline{H}_n(x, m, k)}{1 - x}(-1)^n. \tag{24}
$$

We now seek a recursion formula for this function. We have

$$
\begin{aligned}
&\frac{y^{-m}}{x}\left(y^k + xy^{k+m} + \cdots + (x^{p-1}y^{k+(p-1)m})\right) \\
&= \frac{1}{x} y^{k-m} + (y^k + xy^{k+m} + \cdots + x^{p-1}y^{k+(p-1)m}) - x^{p-1}y^{k+(p-1)m}.
\end{aligned} \tag{25}
$$

---

(9) This result was obtained by Dr. A. M. Mood, following a suggestion by the writer.

Differentiate this relation n times exponentially and put $y=1$. We obtain

$$(26)\qquad [(f(x, m, k) - m)^n \equiv xf_n(x, m, k) + (k - m)^n(1 - x^p) \pmod{p}.$$

Now (24) with (23) gives

$$(27)\qquad f_n(x, m, k) \equiv (-1)^n \frac{1 - x^p}{1 - x} \overline{H}_n(x, m, k) \pmod{p},$$

which, applied to (26), gives

$$(28)\qquad [(\overline{H} + m)^n = x\overline{H}_n + (m - k)^n(1 - x).$$

The first three H's given by this recursion formula are

$$\overline{H}_1 = \frac{(m - k)(x - 1) + m}{x - 1} = \frac{k(1 - x) + mx}{x - 1},$$

$$\overline{H}_2 = \frac{[(m - k)(x - 1) + m]^2 + m^2x}{(x - 1)^2},$$

$$\overline{H}_3 = \frac{[(m - k)(x - 1) + m]^3 + m^2x^2(4m - 3k) + m^2x(m + 3k)}{(x - 1)^3}.$$

Now (28) is not a direct extension of (12). To obtain such an extension of the latter formula we use

$$y^k + xy^{k+m} + \cdots + x^{p-1}y^{k+(p-1)m} - y^k + x^py^ky^{pm} = \frac{y^{k+m}x(1 - y^{pm}x^p)}{1 - xy^m},$$

which, differentiated n times with $y=1$, gives

$$(29)\qquad f_n(x, m, k) - k^n(1 - x^p) \equiv x(1 - x^p) \left[\frac{\overline{d^n}}{dy^n}\left(\frac{y^{k+m}}{1 - xy^m}\right)\right]_{y=1} \pmod{p}.$$

Putting

$$(30)\qquad H_n(x, m, k) = (-1)^n(1 - x) \left[\frac{\overline{d^n}}{dy^n}\left(\frac{y^{k+m}}{1 - xy^m}\right)\right]_{y=1},$$

we then obtain

$$(31)\qquad -f_n(x, m, k) - k^n(1 - x^p) \equiv (-1)^n x \frac{1 - x^p}{1 - x} H_n(x, m, k) \pmod{p}.$$

Using this in connection with (26), we find

$$(32)\qquad [(H + m)^n = xH_n + (-k)^n(1 - x).$$

For $k=0$, $m=1$, this reduces to (12).

3. **Bernoulli numbers.** Set

$$S_n(p) = 0^n + 1^n + 2^n + \cdots + (p-1)^n; \qquad 0^0 = 1,$$

where p is a prime. If $n < p-1$, it is known that $S_n(p) \equiv 0 \pmod{p}$. We then write under this restriction

$$S_n(p) \equiv pa_n \pmod{p^2}, \tag{33}$$

where a_n is some integer.

Consider

$$(1 + x + x^2 + \cdots + x^{p-1})x = x + x^2 + \cdots + x^p. \tag{34}$$

We now differentiate exponentially with respect to x, using Leibnitz's theorem. We obtain

$$\sum_{i=0}^{n} C_{n,i}xf_i(x) = x + 2^n x^2 + 3^n x^3 + \cdots + p^n x^p. \tag{35}$$

Restrict n to be greater than 1. Now (35) can be written

$$\sum_{i=0}^{n} C_{n,i}xf_i(x) = f_n^{(p)}(x) + p^n x^p. \tag{36}$$

Setting $x=1$, we have

$$\sum_{i=0}^{n} C_{n,i}S_i(p) = S_n(p) + p^n, \tag{37}$$

which, in view of (33) and of the fact that $n>1$, is then

$$\sum_{i=0}^{n} C_{n,i}pa_i \equiv pa_n \pmod{p^2}. \tag{38}$$

Dividing through by p, we may write in symbolic form

$$[(a+1)^n \equiv a_n \pmod{p}. \tag{39}$$

Now consider the recursion formula

$$[(b+1)^n = b_n \tag{40}$$

for $n<p-1$. If we determine $b_1, b_2, \cdots, b_{n-1}$ in turn with the use of this, each denominator of the fractions so obtained will be prime to p. In view of (39) we may also obtain a_n in the same manner. Consequently, we may write

$$S_n(p) \equiv pb_n \pmod{p^2}. \tag{41}$$

It is known that $S_n(p) \equiv 0 \pmod{p^2}$ if n is odd, greater than 1, and less than $p-1$. Consequently,

$$pb_n \equiv 0 \pmod{p^2}, \qquad b_n \equiv 0 \pmod{p}$$

for any $p > n+1$; hence.

$$b_n = 0, \qquad \text{for } n \text{ odd and greater than } 1.$$

4. We have, obviously,

$$\frac{x^p - 1}{x - 1} x^k = \frac{x^{(p+k)} - 1}{x - 1} - \frac{x^k - 1}{x - 1}, \tag{42}$$

which may be written

$$\begin{aligned}(1 + x + x^2 + \cdots + x^{p-1})x^k = 1 + x + \cdots + x^{p-1} \\ + x^p(1 + x + \cdots + x^{k-1}) - (1 + x + \cdots + x^{k-1}).\end{aligned} \tag{43}$$

Differentiating this relation exponentially n times with respect to x and collecting the terms whose coefficients are divisible by p^2, we obtain

$$[(f(x) + k)^n = f_n(x) + x^p f_n^{(k)}(x) + npx^p f_{n-1}^{(k)}(x) + p^2 P(x) - f_n^{(k)}(x), \tag{44}$$

where $P(x)$ is a polynomial in x with integral coefficients. Put $x = 1$; we then obtain

$$[(S(p) + k)^n = S_n(p) + npS_{n-1}(k) + p^2 W, \tag{45}$$

where W is an integer. We employ (41) in connection with (45) for each S occurring in the expansion of the left-hand member and for each S occurring in the right-hand member; after dividing the resulting expression through by p, we obtain the relation

$$[(b + k)^n \equiv b_n + nS_{n-1}(k) \pmod{p}. \tag{46}$$

As none of the terms in this congruence depends on p, we obtain the Bernoulli summation formula

$$[(b + k)^n - b_n = nS_{n-1}(k). \tag{47}$$

With (46) as a base we shall now prove the formula[6, 10]

$$[(mb + sm + k)^n - [(mb + k)^n = nm \sum_{i=0}^{s-1} (k + im)^{n-1}. \tag{48}$$

Consider

$$\begin{aligned}(1 + x^m + \cdots + x^{m(p-1)})x^{sm+k} - (1 + x^m + \cdots + x^{m(p-1)})x^k \\ = x^{pm}(x^k + x^{k+m} + \cdots + x^{k+(s-1)m}) \\ - (x^k + x^{k+m} + \cdots + x^{k+(s-1)m}),\end{aligned} \tag{49}$$

(10) Due to Glaisher, Quarterly Journal of Mathematics, vol. 31 (1900), pp. 193–199. For another proof by the writer see American Mathematical Monthly, vol. 36 (1929), pp. 36–37.

where the k and m can be in any ring including the rational field. Employing the concept of generalized exponents used by the writer (Vandiver([3])) the exponential differentiation with respect to x gives

$$(50)\qquad \begin{aligned}&[(f(x^m) + (sm + k))^n - [(f(x^m) + k)^n \\ &\qquad = npmx^{pm}(k^{n-1}x^k + (k + m)^{n-1}x^{k+m} + \cdots \\ &\qquad + (k + (s - 1)m^{n-1})x^m) + p^2P(x),\end{aligned}$$

from which we easily obtain (48), after using (41).

Various arithmetical results not involving explicitly the Bernoulli numbers may be derived by the exponential methods we have been employing. For example, if a is an integer such that

$$a^d \equiv 1 \pmod{p^\alpha},$$

then

$$(51)\qquad (xa^d - 1)^n = \sum_{i=0}^{n} C_{n,i}x^i(-1)^i a^{id}.$$

Differentiating exponentially, with respect to x, k times for $k<m$, we find

$$(52)\qquad P(x)(xa^d - 1)^{n-k} = \sum_{i=0}^{n} C_{n,i}x^i i^k(-1)^i a^{id},$$

where P is a polynomial in x. Put $x=1$. Then we obtain

$$(53)\qquad P(a^d - 1)^{n-k} = \sum_{i=0}^{n} C_{n,i}i^k(-1)^i a^{id},$$

whence

$$(54)\qquad \sum_{i=0}^{n} C_{n,i}i^k(-1)^i a^{id} \equiv 0 \pmod{p^{\alpha(n-k)}}$$

for $k<n$. Putting

$$d = p - 1, \qquad \alpha = 1, \qquad n = 1,$$

we have Fermat's theorem.

5. **Generalized Bernoulli numbers of the first order.** In (48) set $s=p$, a prime. Then expand the left-hand member according to powers of p. The result can be expressed in the form

$$(55)\qquad \begin{aligned}S_{n-1}(m, k, p) = C_{n,1}\Big[&(mb + k)^{n-1}\frac{p}{n} + C_{n,2}\Big[(mb + k)^{n-2}\frac{p^2m}{n} + \cdots \\ &+ \frac{p^rm^{p-1}}{r + 1}C_{n-1,r}p\Big[(mb + k)^{n-1-r} + \cdots + \frac{p^n}{n}.\end{aligned}$$

From this we may show that for $p>2$ and $(mb+k)^0=1$

$$S_n(m,\,k,\,p) \equiv [(mb+k)^n p \pmod{p}. \tag{56}$$

To obtain a proof by induction assume that in (55) for $i<n-1$, $[(mb+k)^i$ may be expressed as a fraction with the denominator prime to p. Also for $r>0$ and p odd we have

$$\frac{p^r}{r+1} \equiv 0 \pmod{p}, \tag{57}$$

since $p^r>r+1$ for p odd, as is seen from

$$p^r \geqq (1+2)^r \geqq 1+2r.$$

Hence (55) gives (56). For p even and $n=2$, (56) also holds, since we may verify that

$$2(mb+k)^2 \equiv S_2(m,\,k,\,2) \pmod{2}. \tag{58}$$

For brevity set

$$h_n = [(mb+k)^n.$$

Now for n odd, p odd, and m prime to p, the expression on the left of (56) is divisible by p, for n is not a multiple of $\overline{p-1}$ since $\overline{p-1}$ is even, and for $m\equiv 0 \pmod{p}$, $S_n(m,\,k,\,p)$ is obviously $\equiv 0 \pmod{p}$. Hence h_n does not have p as a factor of its denominator, except possibly when $p=2$.

Now for $p=2$ (57) holds for $r>1$, so that

$$2^r = (1+1)^r > 1+r$$

for $r>1$. Hence (55) gives for $p=2$, $n>1$ and odd,

$$\sum_{s=0}^{1} (sm+k)^n \equiv 2h_n + \frac{m}{n+1} C_{n+1,2}h_{n-1}2^2 \pmod{2}, \tag{59}$$

$$k^n + (m+k)^n \equiv 2h_n + mn(2h_{n-1}) \pmod{2}.$$

Since $k^n\equiv k \pmod{2}$, for m odd we have

$$1 \equiv 2h_n + (2h_{n-1}) \pmod{2}, \tag{60}$$

and for m odd, n even, we have

$$1 \equiv 2h_n \pmod{2}. \tag{60a}$$

Now from this we cannot have both h_n and h_{n-1} with 2 in the denominator for n odd. Also for m even 2 will not appear in the denominator for n either odd or even. Now since 2 is in the denominator of h_2 for m odd by (58), h_3 is integral from (60). Hence (60) and (60a) give h_n integral for n odd and greater than 1.

Consequently,

$$[(bm + k)^n$$

is integral except when $n=1$ with m odd, and also

$$S_n(m, k, 2) \equiv h_n \cdot 2 \pmod 2 \tag{61}$$

for n even.

Consider

$$h_n - \sum \frac{S_n(m, k, p)}{p}, \tag{62}$$

where the p's range over all the distinct primes less than or equal to $n+1$. For a particular p, say p', of this type the fraction

$$b_n - \frac{S_n(m, k, p')}{p'}$$

may be expressed with a denominator prime to p' by (56) and (61), and the remainder of the expression in (62) obviously has the same property. Hence (62) must have no primes in its denominator and is therefore an integer. Also,

$$S_n(m, k, p) \equiv 0 \pmod p$$

for m prime to p and $n \not\equiv 0 \pmod{\overline{p-1}}$, and obviously also holds for $m \equiv 0 \pmod p$. For $n \equiv 0 \pmod{p-1}$ and m prime to p

$$S_n(m, k, p) \equiv p - 1 \equiv -1 \pmod p,$$

whence the theorem follows[11].

THEOREM I. *If m and k are integers, then if n is even and greater than* 0,

$$\Big[(mb + k)^n = A_n - \sum_{i=1}^{s} \frac{1}{p_i},$$

where $p_1, p_2, \cdots, p_s$ are the distinct primes which are prime to m and such that $n \equiv 0 \pmod{p_i - 1}$, A_n being some integer. If n is odd, then $[(mb+k)^n$ is an integer except when $n=1$ with m odd.

6. In another paper[12] the writer obtained the formulas

---

(11) This theorem was first stated by the writer without proof in Proceedings of the National Academy of Sciences, vol. 23 (1937), p. 556, and the present proof was there briefly indicated. Another proof was given in Duke Mathematical Journal, loc. cit., Theorem III. A third proof is given in §7 of the present paper, but the ordinary von Staudt-Clausen theorem is assumed therein.

(12) Vandiver, Annals of Mathematics, (2), vol. 27 (1926), p. 174 (10); p. 175 (13).

$$
\begin{aligned}
(x^m - 1)\sum_{n=0}^{a-1} C_{a,n}k^{a-n}m^{n+1}b_nf_{a-n}(x) \\
\equiv \sum_{\rho} \frac{x(\rho^k - 1)(x^p - 1)(x^m - 1)F_a(\rho)}{p(x - \rho^k)(x - 1)} \\
- axm(x^m - 1)\sum_{\rho}\sum_{l=0}^{k-1} \frac{\rho^l h_{a-1}^{(k,l)}(x)}{x - \rho^k} \pmod{p},
\end{aligned} \tag{63}
$$

and

$$
F_n(\rho) = (n - 1)p\frac{mf_{n-1}(\rho)}{\rho^p - 1} + p^2C(\rho). \tag{64}
$$

Applying the second formula to the first, dividing the first through by x and setting $x = 0$, we find

$$
t_n(m, k) \equiv {\sum_{\rho}}' \frac{(-1)^{n+1}\rho^k f_{n-1}(\rho)}{1 - \rho^p} + \sum_{l=0}^{k-1}\sum_{\rho} l^{n-1}\rho^{l-k} \pmod{p}, \tag{65}
$$

where $\sum_{\rho}'$ indicates summation over all the mth roots of unity except unity, and $\sum_{\rho}$ indicates summation over all the distinct roots of unity; and where

$$
\frac{[(mb + k)^n - b_n}{n} = t_n(m, k). \tag{66}
$$

When $k \geqq m$, $(m, p) = 1$, $m > 1$, the relation (65) may also be given in the form

$$
t_n(m, k) \equiv {\sum}' \frac{(-1)^{n+1}\rho^k f_{n-1}(\rho)}{1 - \rho^p} + m\sum_{s=1}^{[k/m]} (k - sm)^{n-1} \pmod{p}. \tag{67}
$$

This is a companion formula to (47), but the latter is not obtainable[13] from (67), as (67) does not hold for $n = 1$. Using (8) and (13), we find

$$
f_n(x) \equiv x(1 - x)^{p-n-1}R_n(x) \pmod{p}. \tag{68}
$$

Applying this to (67), we find the equality

$$
t_n(m, k) = {\sum_{\rho}}' \frac{(-1)^{n+1}\rho^{k+1}R_{n-1}(\rho)}{(1 - \rho)^n} + m\sum_{s=1}^{[k/m]} (k - sm)^{n-1}. \tag{69}
$$

In the relation (65) suppose that $m > k$. Then the last term on the right becomes zero, so that we obtain[14]

$$
t_n(m, k) = {\sum_{\rho}}' \frac{(-1)^{n+1}\rho^{k+1}R_{n-1}(\rho)}{(1 - \rho)^n} = -{\sum_{\rho}}' \frac{\rho^{k+1}R_{n-1}(\rho)}{(\rho - 1)^n}, \tag{70}
$$

---

(13) Relations (65) and (67) were given without proof in another paper by the writer, Proceedings of the National Academy of Sciences, vol. 25 (1939), p. 200.

(14) Frobenius, loc. cit., p. 827.

7. Application of some relations due to Frobenius(14). We have

$$\frac{m}{\rho - 1} = \sum_{s=1}^{m-1} s\rho^s, \tag{71}$$

and from (69)

$$\begin{aligned} m^n t_n(m, k) &= -\sum{}'\left(\frac{m}{\rho - 1}\right)^n \rho^{k+1}R_{n-1}(\rho) + I \\ &= -\sum{}'\left(\sum_{s=1}^{m-1} s\rho^s\right)^n \rho^{k+1}R_{n-1}(\rho) + I, \end{aligned} \tag{72}$$

where I is an integer. The expression on the right is a polynomial in ρ with integral coefficients and when summed over all roots of unity except 1 will be an integer, as $\rho+\rho^2+\cdots+\rho^{m-1}=-1$. Hence $m^n t_n(m, k)$ is an integer.

To show that $m((mb+k)^n-b_n)$ is an integer we note that

$$\frac{m^n}{n}\left[[(mb + k)^n - b^n\right]$$

is an integer; expanding this, we obtain

$$\frac{m^n}{n}\left[(m^n - 1)b_n + k^n + m(R)\right] = \frac{m^n}{n} K. \tag{73}$$

Since (73) is an integer, the fractions with denominators prime to m will cancel out and we shall be concerned only with primes in the denominators which divide m. By the von Staudt-Clausen theorem these appear only to the first power, hence in (73) $m(R)$ is an integer.

Let d denote the denominator of b_n; then

$$\frac{m'}{n}(K) \tag{74}$$

is an integer in which $m'=(m^n, nd)$. But mn is divisible by m', since prime factors of d occur only to the first power; hence we can replace m' by mn in (74) and obtain the desired result.

We shall now show that $[(mb+k)^n$ is an integer for n odd and greater than 1 as in Theorem I. First note that $m[(mb+k)^n$ is an integer, since b_n is zero for n odd and greater than 1. Now if $[(mb+k)^n$ were a fraction, it would have a denominator which divided m. Hence, since every term of $[(mb+k)^n$ contains m except the final term k^n, we see that $[(mb+k)^n$ is an integer, except in the case when $n=1$ and m is odd. In the latter case we have $mb+k=-(1/2)m+k$.

8. We shall now give another proof of our Theorem I, using the von Staudt-Clausen theorem in the expansion of $[(mb+k)^n$, that is,

$$[(mb+k)^n = m^n b_n + n m^{n-1} b_{n-1} k + \cdots + k^n, \tag{75}$$

and obtaining

$$\Big[(mb+k)^n = I + \sum \frac{m_i}{r_i}, \tag{76}$$

where the m_i are integers less than r_i. Now

$$m\big[\big[(mb+k)^n - b_n\big] = I_i, \tag{77}$$

where I_i represents an integer. Hence,

$$mI + m\sum \frac{m_i}{r_i} - m\alpha_n + m\sum \frac{1}{q_t} = I_1, \tag{78}$$

or

$$m\left(\sum \frac{m_i}{r_i} + \sum \frac{1}{q_t}\right) = I_2, \tag{79}$$

where I_2 represents an integer which combines mI, $m\alpha_n$, and terms wherein q_t and r_i divide m. Now in (79) no q_t or r_i divides m, hence

$$\left(\sum \frac{m_i}{r_i} + \sum \frac{1}{q_t}\right)$$

must be an integer I_3; therefore no r_i is different from some q_t and vice versa, and hence

$$\sum \frac{m_i + 1}{q_t} = I_3.$$

Since each term in this sum must be an integer and since $|m_i| < q_t$, we obtain $m_i = -1$. Thus, we have proved Theorem I.

9. **Illustration of our congruence methods.** In order to illustrate our congruence methods further we show how a known property of the generalized b's may be easily derived. By direct expansion we find that

$$\sum_{s=0}^{p-1} (k+sm)^n = \sum_{i=0}^{n} C_{n,i} s_i(p) m k^{n-i}.$$

Now let $n+1 < p$; using (41) we find that

$$p\Big[(mb+k)^n \equiv \sum_{s=0}^{p-1} (k+sm)^n \pmod{p^2}, \tag{80}$$

whence

$$p\sum_{i=0}^{m-1}\Big[(mb+i)^n = \sum_{i=0}^{m-1}\sum_{s=0}^{p-1}(ms+i)^n = \sum_{i=0}^{mp-1} i^n = S_n(mp) \equiv mpb_n \pmod{p^2}.$$

Consequently[15],

$$\sum_{i=0}^{m-1} [(mb+i)^n = mb_n, \tag{81}$$

10. We now develop some congruences of a novel character involving the Mirimanoff polynomials, and leading to congruences of a new type relating to the Bernoulli numbers. Obviously,

$$(x^k-1)\frac{x^p-1}{x-1} = (x^p-1)\frac{x^k-1}{x-1},$$

whence, for p an odd prime,

$$\begin{aligned}(x^k-1)f_{p-1}(x) - kx^kf_{p-2}(x) + k^2x^kf_{p-3}(x) - \cdots + x^kk^{p-1}f_0(x)& \\ \equiv (x^p-1)\sum_{l=1}^{k-1} l^{p-1}x^l \pmod{p}&,\end{aligned}$$

or, if $[u]$ is the greatest integer in u,

$$\begin{aligned}&-kf_{p-2}(x) + k^2f_{p-3}(x) + \cdots + k^{p-1}f_0(x) \\ &\quad\equiv (x^p-1)\sum_{l=1}^{k-1} l^{p-1}x^{l-k} + (x^{-k}-1)\frac{x-x^p}{1-x} \pmod{p} \\ &\quad\equiv \frac{(x^{p-k}-x^{-k})(x-x^k)}{1-x} + \frac{(x^{-k}-1)(x-x^p)}{1-x} \\ &\qquad - (x^p-1)(x^{p-k}+x^{2p-k}+\cdots+x^{p[k/p]-k}) \pmod{p}, \\ &\quad\equiv 1-x^{p-k}-(x^p-1)(x^{p-k}+x^{2p-k}+\cdots+x^{p[k/p]-k}) \pmod{p},\end{aligned}$$

where we understand that the second member is zero if $[k/p]=0$. Assume $p>n+1$, replace k by kr and multiply by r^{p-1-n}, let r take on $1, 2, \cdots, p-1$, and add. We have, after dividing by $(1-x^p)$

$$\begin{aligned}\frac{(-1)^{n+1}k^nf_{p-n-1}(x)}{1-x^p} &\equiv \sum_{r=1}^{p-1}\frac{1-x^{p-kr}}{1-x^p}(r)^{p-1-n} \\ &\quad + \sum r^{p-1-n}(x^{p-kr}+\cdots+x^{p[kr/p]-kr}) \\ &\equiv -\frac{x^pf_{p-n-1}(x^{-k})}{1-x^p} \\ &\quad + \sum_{r=1}^{p-1} r^{p-1-n}(x^{p-kr}+\cdots+x^{p[kr/p]-kr}).\end{aligned} \tag{82}$$

---

(15) Kummer, Crelle's Journal, vol. 40 (1850), pp. 119–121; Blissard, Quarterly Journal of Mathematics, vol. 4 (1861), p. 288. There are also later references.

Let $x=\rho$, an mth root of unity different from 1, $(m, p)=1$ and let $\sum'$ indicate summation over all ρ's$\neq 1$, and set $k=2$. We obtain, modulo p,

$$
\begin{aligned}
\sum{}' \frac{(-1)^{n+1} f_{p-n-1}(\rho) k^n}{1-\rho^p} &\equiv \sum{}' f_{p-n-1}(\rho^{-k}) \\
&\quad - \sum{}' \frac{f_{p-n-1}(\rho^{-k})}{1-\rho^p} + \sum{}' \sum_{s=(p+1)/2}^{p-1} s^{p-1-n}\rho^{p-sk}.
\end{aligned} \tag{83}
$$

Let $m<(p-1)/2$ and $(m, 2)=1$, $p>2$ and $m>2$, and consider the expression

$$\sum_{s=(p+1)/2}^{p-1} s^{p-1-n}\rho^{p-sk},$$

and we shall determine the terms in which the exponents of ρ are $\equiv 0 \pmod m$. If l is one such, then

$$p - lk \equiv 0 \pmod m,$$

where l is in the set $(p+1)/2, \cdots, p-1$. From the above congruence $(p+m)/2$ is a solution, but $(p-m)/2$, although it satisfies the congruence, is not in the set mentioned. Hence the solutions we need for $[p/m]$ odd are:

$$\frac{p+m}{2}, \frac{p+3m}{2}, \cdots, \frac{p+[p/m]m}{2}.$$

For $[p/m]$ even, it is replaced by $([p/m]-1)$ in the above.

$$\sum{}' \sum_{s=(p+1)/2}^{p-1} s^{p-1-n}\cdot\rho^{p-sk} \equiv - \sum_{s=(p+1)/2}^{p-1} s^{p-1-n} + m\sum_{v=1}^{h}\left(\frac{p+vm}{2}\right)^{p-1-n}, \tag{84}$$

where $h=[p/m]$ or $[p/m]-1$ according as $[p/m]$ is odd or even.

Now

$$\frac{p+jm}{2} \equiv \frac{jm}{2} \pmod p,$$

so that the last term on the right is congruent, modulo p, to

$$m^{p-n}\left(\left(\frac{p+1}{2}\right)^{p-1-n} + \left(\frac{p+3}{2}\right)^{p-1-n} + \cdots + \left(\frac{p+h}{2}\right)^{p-1-n}\right),$$

where h is defined as above.

If in (83) we employ (67) where the k used in the latter congruence equals zero, we obtain new relations involving Bernoulli numbers in view of (66).

11. **Bernoulli numbers of the second order.** We employ the identity[16]

[16] Vandiver, Annals of Mathematics, (2), vol. 29 (1928), p. 171.

$$\frac{x^j - y^k}{(x-1)(y-1)} = \frac{\sum_{n=0}^{j-1} x^n y^{k-a_n}}{y-1} - \frac{\sum_{l=0}^{k-1} y^l x^{j-b_l}}{x-1}, \tag{85}$$

where $a_n = [nk/j]$, $b_l = [lj/k]$, $[u]$ is the greatest integer in u, j and k are positive integers, x and y arbitrary. Multiply by $(x^p-1)(y^p-1)$, where p is a prime greater than n, set $x = xz^k$, $y = yz^j$, divide by z^{jk}, differentiate exponentially a times with respect to z and set $z = 1$. We then have, reducing the right-hand member, modulo p^3, and where $h_a^{(m,k)}(x)$ is the $f_a(x, m, k)$ of (20),

$$\begin{aligned}(x^j - y^k)\Big[(kf(x) + jf(y))^a \equiv{}& y^k(x^p - 1)\sum_{n=0}^{j-1} x^n y^{-a_n} h_a^{(j,c_n)}(y) \\ &+ \sum_{n=0}^{j-1} akx^{p+n} y^{k-a_n} p h_{a-1}^{(j,c_n)}(y) \\ &+ C_{a,2}k^2 x^p y^{k-a_n} \sum_{n=0}^{j-1} p^2 h_{a-2}^{(j,c_n)}(y) \\ &- x^j(y^p - 1)\sum_{l=0}^{k-1} y^l x^{-b_l} h_a^{(k,d_l)}(x) \\ &- apj\sum_{l=0}^{k-1} y^{p+l} x^{j-b_l} h_{a-1}^{(k,d_l)}(x) \\ &- C_{a,2}j^2 p^2 \sum_{l=0}^{k-1} h_{a-2}^{(k,d_l)}(x) \pmod{p^3},\end{aligned} \tag{86}$$

where c_n denotes the least positive or zero residue of nk, modulo j, while d_l denotes the least positive or zero residue of lj, modulo k. Now set $x = xz^k$ and differentiate once with respect to z; we then have, modulo p^3, after setting $z = 1$,

$$\begin{aligned}kjx^j\big[(kj(x) + jf(y))^a &+ (x^j - y^k)\big[\overline{D_z}[(kf(x) + jf(y))^a]_{z=1} \\ &\equiv y^k(x^p - 1)\Big[\overline{D_z}\Big(\sum_{n=0}^{j-1} x^n z^{nk} y^{-a_n} h_a^{(j,c_n)}(y)\Big)\Big]_{z=1} \\ &+ pkx^p y^k \sum_{n=0}^{j-1} x^n y^{-a_n} h_a^{(j,c_n)}(y) + \sum_{n=0}^{j-1} ak^2(p+n) y^{k-a_n} p h_{a-1}^{(j,c_n)}(y) \\ &- apj\sum_{l=0}^{k-1} y^{p+l} x^{j-b_l} h_a^{(k,d_l)}(x) - apj^2 \sum_{l=0}^{k-1} (k-l) y^{p+l} x^{j-b_l} h_{a-1}^{(k,d_l)}(x)\end{aligned} \tag{87}$$

plus terms of the form $p^2 g(x)$, where each g is divisible by some $h_t(x)$.

Setting $x = y = 1$, dividing through by p^2, and using

$$\begin{aligned}p(rb + s)^n \equiv{}& s^n + (r + s)^n + (2r + s)^n + \cdots \\ &+ ((p-1)r + s)^n \pmod{p^2},\end{aligned} \tag{88}$$

which follows from (55) for $p>n$, we find, since the result is independent of p,

$$(89)\qquad \begin{aligned} kj\Big[(kb+jb')^a &= k\sum_{n=0}^{j-1}\Big[(jb+c_n)^a + a\sum_{n=0}^{j-1}k^2n\Big[(jb+c_n)^{a-1} \\ &\quad - aj\sum_{l=0}^{k-1}\Big[(kb+d_l)^a - aj^2\sum_{l=0}^{k-1}\Big[(kb+d_l)^{a-1}(k-l), \end{aligned}$$

or since by (81)

$$(90)\qquad \sum_{n=0}^{j-1}\Big[(jb+c_n)^a = jb_a; \qquad \sum_{l=0}^{k-1}\Big[(kb+d_l)^a = kb_a,$$

we may write

$$(91)\qquad \begin{aligned} kj\Big[(kb+jb')^a &= jk(1-a)b_a - ak^2j^2b_{a-1} + a\sum_{n=0}^{j-1}k^2n\Big[(jb+c_n)^{a-1} \\ &\quad + a\sum_{l=0}^{k-1}j^2l\Big[(kb+d_l)^{a-1}. \end{aligned}$$

For $h=2, j=1$, we find[17]

$$(92)\qquad 2\big[(2b+b')^a = 2(1-a)b_a - 4ab_{a-1} + a(2b+1)^{a-1}.$$

Now since $(-b)^n = b^n$ except for $n=1$, we have

$$(93)\qquad \big[(kb-jb')^a = \big[(kb+jb')^a + ajk^{a-1}b_{a-1},$$

whence from (91)

$$(94)\qquad \begin{aligned} kj\Big[(kb-jb')^a &= jk(1-a)b_a - ak^2j^2b_{a-1} + ajk^{a-1}b_{a-1} \\ &\quad + a\sum_{n=0}^{j-1}k^2n\Big[(jb+c_n)^{a-1} + a\sum_{l=0}^{k-1}j^2l\Big[(kb+d_l)^{a-1}. \end{aligned}$$

Setting $j=k=1$ in (91) and (94), we obtain the well known relations

$$(95)\qquad \big[(b+b')^a = (1-a)b_a - ab_{a-1},$$

and

$$(96)\qquad \big[(b-b')^a = (1-a)b_a.$$

Now for a odd we have

$$(97)\qquad \big[(kb+jb')^a = akb_1(jb)^{a-1} + a(kb)^{a-1}jb_1,$$

so that from (91) we have the formula

---

(17) Bell, these Transactions, vol. 24 (1922), p. 106.

$$
\begin{aligned}
ak^2b_1b_{a-1}j^a + aj^2b_1b_{a-1}k^a &= ab_1b_{a-1}(k^2j^a + j^2k^a)\\
&= -ak^2j^2b_{a-1} + a\sum_{n=0}^{j-1} k^2n\left[(jb + c_n)^{a-1}\right.\\
&\quad + a\sum_{l=0}^{k-1} j^2l\left[(kb + d_l)^{a-1}\right..
\end{aligned}
\tag{98}
$$

There is an analogous relation from (94).

12. If h is any integer, we have from (91)

$$
\begin{aligned}
\left[(kb + jb' + h)^n\right. &= \sum_{a=0}^{n} C_{n,a}\left[(kb + jb')^a h^{n-a}\right.\\
&= jk\sum_{a=0}^{n} C_{n,a}b_a h^{n-a} - jk\sum_{a=0}^{n} C_{n,a}ab_a h^{n-a}\\
&\quad - k^2j^2\sum_{a=1}^{n} aC_{n,a}h^{n-a}b_{a-1} + \sum_{a=1}^{n}\sum_{i=0}^{j-1} k^2iaC_{n,a}\left[(jb + c_i)^{a-1}h^{n-a}\right.\\
&\quad + \sum_{a=1}^{n}\sum_{i=0}^{k-1} j^2laC_{n,a}\left[(kb + d_l)^{a-1}\right.\\
&= jk\left[(b + h)^n\right. - jkn\sum_{a=1}^{n} C_{n-1,a-1}h^{n-a}b_a\\
&\quad - k^2j^2n\sum_{a=1}^{n} C_{n-1,a-1}h^{n-a}b_{a-1}\\
&\quad + \sum_{a=1}^{n}\sum_{i=1}^{j-1} k^2inC_{n-1,a-1}\left[(jb + c_i)^{a-1}h^{n-a}\right.\\
&\quad + \sum_{a=1}^{n}\sum_{l=0}^{k-1} j^2lnC_{n-1,a-1}\left[(kb + d_l)^{a-1}h^{n-a}\right.,
\end{aligned}
\tag{99}
$$

where we have changed the notation employed in (91). n now takes the place of a and c_i denotes the least positive or zero residue of ik, modulo j. Now

$$
\sum_{a=1}^{n} C_{n-1,a-1}b_a h^{n-a} = \left[b(h + b)^{n-1}\right.,
\tag{100}
$$

and

$$
\left[b(h + b)^{n-1}\right. = \left[(h + b)^n\right. - h\left[(h + b)^{n-1}\right.,
\tag{101}
$$

and we have this theorem:

Theorem II. *If h, j and k are integers, $k>0$, $j>0$; c_i represents the least positive or zero residue of ik* (mod j); *d_l represents the least positive or zero residue of lj*, (mod k); *$n>0$; then*

$$
\begin{aligned}
jk\big[(kb + jb' + h)^n = jk\big[(b + h)^n &- jkn\big[(b + h)^n \\
&+ jknh\big[(b + h)^{n-1} - k^2j^2n\big[(b + h)^{n-1} \\
&+ k^2n\sum_{i=0}^{j-1} i\Big[(jb + c_i + h)^{n-1} \qquad (102)\\
&+ j^2n\sum_{l=0}^{k-1} l\Big[(kb + d_l + h)^{n-1}.
\end{aligned}
$$

Now since $(-b)^a = b^a$ except for $a = 1$, we have

$$(103) \qquad \big[(kb - jb' + h)^n = \big[(kb + jb' + h)^n + nj\big[(kb + h)^{n-1}.$$

This formula with (102) gives the following theorem:

THEOREM III. *If h, j and k are integers, n even and greater than* 0, *then*

$$(104) \qquad jk\Big[(kb + jb' + h)^n = I - (1 - n)jk\sum_{i=1}^{s} \frac{1}{p_i},$$

where the p's are the distinct primes such that $n \equiv 0 \pmod{p_i - 1}$, I being some integer.

This is an analogue of the generalized von Staudt-Clausen theorem (I).

Now for n odd, (102) and (103) give

THEOREM IV. *If $(n+1)$ is odd, then for n greater than zero*

$$(105) \qquad jk\Big[(kb + jb' + h)^{n+1} = I_1 + (n + 1)jk\left(\frac{j}{2} + \frac{k}{2} - h\right)\sum_{i=1}^{s} \frac{1}{p_i},$$

where I_1 is some integer, k, j and h are integers, kj is odd and the p's are defined as in Theorem III.

It is not clear that this result is an analogue of any theorem involving the Bernoulli numbers of the first order.

We also have from (102) and (103) the

COROLLARY I. *Any Bernoulli number of the second order can be expressed as a linear combination of Bernoulli numbers of the first order with coefficients whose denominators divide the integers occurring in the original number.*

COROLLARY II. *The expression*

$$(106) \qquad jk\big[(kb + jb' + h)^n + jk(n - 1)\big[(b + h)^n$$

is an integer if n is even, j, k and h integers.

The relation (102) gives, employing Theorem II of another paper([18]) the congruence

([18]) Vandiver, Duke Mathematical Journal, loc. cit.

$$jk\Big[(kb + jb' + h)^n \equiv 0 \pmod{p^a}, \tag{107}$$

where n is odd, p prime, $n-1=p^a r$, $(r, p)=1$, $(p, p_i)=1$, p_i being any von Staudt-Clausen prime of order $n-1$. If n is even, this relation also holds if $n\equiv 0 \pmod{p^a}$, $(p, q)=1$, where q is any von Staudt-Clausen prime of order n.

13. We shall now prove that

$$b_{a-1}(j^2 - j^a) \equiv 2\sum_{i=0}^{j-1} i\Big[(jb + c_i)^{a-1}, \tag{108}$$

with a odd and c_i denoting the least positive or zero residue of ik (mod j). Assuming (108) true for all values of k such that $k<j$, $(k, j)=1$, we have from (91)

$$\begin{aligned} &-b_{a-1}(k^2j^a + j^2k^a) + 2k^2j^2b_{a-1} \\ &\qquad = 2\sum_{i=0}^{j-1} k^2 i\Big[(jb + c_i)^{a-1} + 2\sum_{l=0}^{k-1} j^2 l\Big[(kb + d_l)^{a-1}, \end{aligned} \tag{109}$$

and also

$$-b_{a-1}(k^2j^a + j^2k^a - 2k^2j^2) = 2\sum_{i=0}^{j-1} k^2 i\Big[(jb + c_i)^{a-1} + j^2(k^2 - k^a)b_{a-1}, \tag{110}$$

using $c_i = c_r$ where $d_l = d_r$ with $r\equiv l \pmod k$, $(r, k)=1$, $r<k$; whence, the result.

14. Let $\rho^j = 1$; then

$$\Big[(jb + c_n)^a \equiv b_a + (-1)^{a+1}a\sum{}' \frac{\rho^{c_n}f_{a-1}(\rho)}{1-\rho^p} \pmod p, \qquad a>1; \tag{111}$$

where $\sum'$ indicates summation over all distinct jth roots of unity different from 1; then

$$\begin{aligned} \frac{1}{2}b_a(j^2 - j^{a+1}) &\equiv \sum_{n=1}^{j-1} n\Big[(jb + c_n)^a \\ &\equiv \sum_{n=1}^{j-1} nb_a + (-1)^{a+1}a\sum_{n=1}^{j-1}\sum{}' \frac{n\rho^{c_n}f_{a-1}(\rho)}{1-\rho^p} \pmod p, \end{aligned} \tag{112}$$

or

$$\frac{j - j^{a+1}}{2}\, b_a \equiv (-1)^{a+1}a\sum_{n=1}^{j-1}\sum{}' \frac{n\rho^{c_n}f_{a-1}(\rho)}{1-\rho^p}, \tag{113}$$

$$\frac{(1 - j^a)b_a}{2a} \equiv \sum{}' \frac{f_{a-1}(\rho)}{(1-\rho^k)(1-\rho^p)}, \quad a \text{ even and greater than } 1. \tag{114}$$

The latter relation may be proved directly by noting that if $(k, j)=1$,

$$\frac{f_{a-1}(\rho)}{(1-\rho^k)(1-\rho^p)} = \frac{\rho^{-k}(-1)^{a-1}f_{a-1}(\rho^{-1})}{(\rho^k-1)(\rho^{-p}-1)}. \tag{115}$$

Hence

$$\sum{}' \frac{f_{a-1}(\rho)}{(1-\rho^k)(1-\rho^p)} = \sum{}' \frac{-\rho^k f_{a-1}(\rho)}{(1-\rho^k)(1-\rho^p)}, \tag{116}$$

and

$$\begin{aligned} &\sum{}'\left(\frac{f_{a-1}(\rho)}{(1-\rho^k)(1-\rho^p)} + \frac{-\rho^k f_{a-1}(\rho)}{(1-\rho^k)(1-\rho^p)}\right) \\ &\qquad = \sum{}' \frac{f_{a-1}(\rho)}{1-\rho^p} \equiv (-1)^a \frac{(1-j^a)b_a}{a} \pmod{p}. \end{aligned} \tag{117}$$

15. Take the obvious identity

$$\frac{x-y}{(x-1)(y-1)} = \frac{y}{y-1} - \frac{x}{x-1}$$

((85) reduces to this for $k=1, j=1$). We may then write

$$(x-y)f_0^{(kp)}(x)f_0^{(jp)}(y) = y(x^{kp}-1)f_0^{(jp)}(y) - x(y^{jp}-1)f_0^{(kp)}(x). \tag{118}$$

Set $x=xz$, $y=yz$, divide by z, then differentiate each member exponentially a times with respect to z, and reduce the terms, modulo p^2; after setting $z=1$, we find

$$\begin{aligned} (x-y)[(f(x)+f(y))^a \equiv{} & (x^{kp}-1)yf_a^{(jp)}(y) + akpx^{kp}yf_{a-1}^{(jp)}(y) \\ & + C_{n,2}k^2p^2x^{kp}yf_{a-2}^{(jp)}(y) - (y^{jp}-1)xf_a^{(kp)}(x) \\ & - ajpy^{jp}xf_{a-1}^{(kp)}(x) \\ & - C_{n,2}j^2p^2y^{jp}xf_{a-2}^{(kp)}(x) \pmod{p^3}. \end{aligned} \tag{119}$$

Set $x=\zeta$, a kth root of unity, and $y=\rho$, a jth root of unity, let p be a prime greater than a, and then sum each member over all values of ζ and ρ except when ζ and ρ are simultaneously 1. We obtain

$$\sum_{\rho,\zeta}\left[\frac{1}{p^2}(f^{(kp)}(\zeta)+f^{(jp)}(\rho))^a \equiv \sum_{\rho,\zeta}\left\{\frac{akp\rho f_{a-1}^{(jp)}(\rho)}{p(\zeta-\rho)} + \frac{aj\zeta f_{a-1}^{(kp)}(\zeta)}{p(\zeta-\rho)}\right\} \pmod{p}.\right. \tag{120}$$

We note that this is symmetric in j and k, hence we need consider only the first term on the right (call it T_1) and obtain T_2, the second term, by interchanging j and k. First, in T_1 we sum $1/(\zeta-\rho)$ with respect to ζ for $\rho\neq 1$. To effect this we employ the identity

$$x^k - 1 = (x - 1)(x - \zeta) \cdots (x - \zeta^{k-1}),$$

where ζ now represents a primitive kth root of unity. Ordinary differentiation of this with respect to x gives

$$kx^{k-1} = \frac{x^k - 1}{x - 1} + \frac{x^k - 1}{x - \zeta} + \cdots + \frac{x^k - 1}{x - \zeta^{k-1}}.$$

Dividing by $x^k - 1$; putting $x = \rho$; and changing signs, we have

$$\frac{k\rho^{k-1}}{1 - \rho^k} = \frac{1}{1 - \rho} + \frac{1}{\zeta - \rho} + \cdots + \frac{1}{\zeta^{k-1} - \rho},$$

which is the desired sum. Hence

$$T_1 = \sum_\rho{}' \frac{ak^2\rho^k f_{a-1}^{(jp)}(\rho)}{p(1 - \rho^k)} + \sum_\zeta{}' \frac{akf_{a-1}^{(jp)}(1)}{p(\zeta - 1)}. \tag{121}$$

Now, using (71), we have

$$\begin{aligned} \sum_\zeta{}' \frac{akf_{a-1}^{(jp)}(1)}{p(\zeta - 1)} &= -\frac{ak(k-1)}{2} \cdot \frac{f_{a-1}^{(jp)}(1)}{p} \\ &= -\frac{1}{2}ak(k-1)\frac{S_{a-1}(jp)}{p} \\ &\equiv -\frac{1}{2}ak(k-1)jb_{a-1} \pmod{p} \end{aligned} \tag{122}$$

for $a < p$. To transform the other term in T_1 we employ the relation, with $n > 0$,

$$x^p(-1)^n f_n\left(\frac{1}{x}\right) \equiv f_n(x) \pmod{p},$$

together with[19]

$$f_n^{(jp)}(\rho) \equiv -njp\frac{f_{n-1}(\rho)}{1 - \rho^p} \pmod{p^2}.$$

In this way we obtain for $a > 2$, modulo p,

$$\begin{aligned} \sum_\rho{}' \frac{\rho^k f_{a-1}(\rho)}{p(1 - \rho^k)} &\equiv \sum_\rho{}' \frac{(a-1)j\rho^k f_{a-2}(\rho)}{(1 - \rho^k)(\rho^p - 1)} \\ &\equiv \sum_\rho{}' \frac{(a-1)j(-1)^{a-2}\rho^p \cdot \rho^k f_{a-2}(\rho^{-1})}{(1 - \rho^k)(\rho^p - 1)} \\ &\equiv \sum_\rho{}' \frac{(a-1)j(-1)^{a-2} f_{a-2}(\rho^{-1})}{(\rho^{-k} - 1)(1 - \rho^{-p})}, \end{aligned}$$

(19) Vandiver, Annals of Mathematics, (2), vol. 27 (1926), p. 175.

that is,

$$(123)\qquad \sum_\rho{}' \frac{\rho^k f_{a-1}(\rho)}{p(1-\rho^k)} \equiv \sum_\rho{}' \frac{(a-1)j(-1)^{a-2}f_{a-2}(\rho)}{(\rho^k-1)(1-\rho^p)} \pmod p.$$

Using (122) and (123) in connection with (121), we obtain

$$T_1 \equiv \sum_\rho{}' \frac{ak^2(a-1)j(-1)^{a-2}f_{a-2}(\rho)}{(\rho^k-1)(1-\rho^p)} - \frac{ak(k-1)}{2} jb_{a-1} \pmod p,$$

and T_2 is obtained from this by interchanging j and k together with ρ and ζ. Add the quantity

$$(124)\qquad \frac{1}{p^2}\left[(f^{(kp)}(1) + f^{(jp)}(1))^a \equiv [(b+b')^a jk \pmod p\right)$$

to both members of (120); after employing

$$\sum_\rho f_n^{(kp)}(\rho) \equiv k^{n+1} b_n p \pmod{p^2}$$

and the corresponding relation involving j, we then obtain (using (95))

$$(125)\qquad \begin{aligned} &\frac{[(kb+jb')^a}{a(a-1)} + \frac{b_{a-1}}{2(a-1)}(k+j) + \frac{b_a}{a} \\ &\equiv k\sum_\rho{}' \frac{(-1)^a f_{a-2}(\rho)}{(\rho^k-1)(1-\rho^p)} + j\sum_\zeta{}' \frac{(-1)^a f_{a-2}(\zeta)}{(\zeta^j-1)(1-\zeta^p)} \pmod p, \end{aligned}$$

where $j>1$, $k>1$, $2<a<p$, with j, k and p prime each to each. Employing (68) for $x=\rho$, $n=a-2$, we obtain the following theorem:

THEOREM V. *If $j>1$, $k>1$, $a>2$, with $(j,k)=1$, ρ a jth root of unity different from 1, ζ a kth root of unity different from 1, then*

$$(126)\qquad \begin{aligned} &\frac{[(kb+jb')^a}{a(a-1)} + \frac{b_{a-1}}{2(a-1)}(k+j) + \frac{b_a}{a} \\ &= k\sum_\rho{}' \frac{\rho R_{a-2}(\rho)}{(1-\rho^k)(\rho-1)^{a-1}} + j\sum_\zeta{}' \frac{\zeta R_{a-2}(\zeta)}{(1-\zeta^j)(\zeta-1)^{a-1}}, \end{aligned}$$

where the summations extend over each distinct value of ρ and ζ, respectively, and the R's are defined as in (13).

We shall now show how to obtain (91) from (125), but subject to the restrictions on the latter relation. Using (71) we have

$$\frac{j}{\rho^k-1} = \sum_{i=1}^{j-1} i\rho^{ik},$$

and letting c_i as before be the least positive or zero residue of ik (mod j), and employing (65), we have

$$(127)\quad \sum_{\rho}{}' \frac{(a-1)j(-1)^{a-2}f_{a-2}(\rho)}{(\rho^k-1)(1-\rho^p)} \equiv \sum_{i=1}^{j-1} i\left[(jb+c_i)^{a-1} - b_{a-1}\frac{j(j-1)}{2}\right. \pmod p,$$

with a similar expression for the other term in the right-hand member of (125); if we substitute in (125) we have (91), after noting that the resulting terms are independent of p.

Using Theorem V, we obtain the corollary:

COROLLARY I. *The expression*

$$k^a j^a\left(\frac{[(kb+jb')^a}{a(a-1)} + \frac{b_{a-1}}{2(a-1)}(k+j) + \frac{b_a}{a}\right)$$

is an integer with the restrictions on j, k and a given in Theorem V.

We also have this corollary:

COROLLARY II. *If j, k and a are restricted as in Theorem* V, *we have, if* $(jk, a(a-1))=1$,

$$2[(kb+jb')^a + ab_{a-1}(k+j) + 2(a-1)b_a \equiv 0 \pmod{a(a-1)}).$$

These results indicate certain analogies between the properties of T_n in (69) and the number expressed by the left-hand member of (125).

16. **Bernoulli numbers of higher order.** Bernoulli numbers of higher order than the second, namely, numbers of the form

$$(128)\quad \begin{aligned} &[(m_r b^{(r)} + m_{r-1}b^{(r-1)} + \cdots + m_1 b' + m_0)^n = b_n(m_r, m_{r-1}, \cdots, m_0), \\ &\qquad\qquad \text{for } r \geqq 3;\ m_i \neq 0;\ i = 1, 2, \cdots, r, \end{aligned}$$

do not have properties as simple as those of the first and second order, since other primes than the von Staudt-Clausen primes appear as factors in the denominators of such numbers. This may be illustrated in the case of the number

$$[(b^{(1)} + b^{(2)} + \cdots + b^{(s)})^n.$$

This may be reduced as follows. Consider first

$$[(b+b'+b'')^n = b_n + nb_{n-1}(b'+b'') + \cdots + C_{n,2}b_{n-2}(b'+b'')^2 + \cdots.$$

Then

$$[(b'+b'')^k = (1-k)b_k - kb_{k-1},$$

so that

$$\Big[(b + b' + b'')^n = \sum_{r=0}^{n} C_{n,r}b'_{n-r}((1 - r)b_r - rb_{r-1})$$

$$= \sum_{r=0}^{n} C_{n,r}b'_{n-r}b_r - \sum_{r=0}^{n} C_{n,r}rb'_{n-r}b_r - \sum_{r=0}^{n} C_{n,r}rb'_{n-r}b_{r-1}. \tag{129}$$

Now

$$\sum_{r=0}^{n} C_{n,r}b'_{n-r}b_r = \Big[(b + b')^n. \tag{130}$$

Also,

$$\sum_{r=0}^{n} rC_{n,r}b'_{n-r}b_r = n\sum_{r=1}^{n} C_{n-1,r-1}b'_{n-r}b_{r-1}b = n\sum\Big[b(b + b')^{r-1}. \tag{131}$$

Now

$$\big[b(b + b')^{r-1} = (1/2)\big[(b + b')(b + b')^{r-1} = (1/2)\big[(b + b')^r,$$

so that

$$\sum_{r=0}^{n} rC_{n,r}b'_{n-r}b_r = \frac{n}{2}\Big[(b + b')^r. \tag{132}$$

Also,

$$\sum_{r=0}^{n} C_{n,r}rb'_{n-r}b_{r-1} = n\sum_{r=1}^{n} C_{n-1,r-1}b'_{n-r}b_{r-1} = n\Big[(b + b')^{r-1}. \tag{133}$$

Hence, using (130), (131), (132), and (133) with (129), we have

$$\Big[(b + b' + b'')^n = \left(1 - \frac{n}{2}\right)\Big[(b + b')^n - n\Big[(b + b')^{n-1}. \tag{134}$$

We shall now prove, by induction on s, the formula ($n > 0$)

$$\begin{aligned}\Big[(b^{(1)}+b^{(2)}+ \cdots +b^{(s)})^n &= \left(1-\frac{n}{s-1}\right)\Big[(b^{(1)}+b^{(2)}+ \cdots +b^{(s-1)})^n \\ &\quad -n\big[(b^{(1)}+b^{(2)}+ \cdots +b^{(s-1)})^{n-1}.\end{aligned} \tag{135}$$

Assume for $s > 2$

$$\begin{aligned}\Big[(b^{(1)}+b^{(2)}+ \cdots +b^{(s-1)})^n &= \left(1-\frac{n}{s-2}\right)\Big[(b^{(1)}+b^{(2)}+ \cdots +b^{(s-2)})^n \\ &\quad -n\big[(b^{(1)}+b^{(2)}+ \cdots +b^{(s-2)})^{n-1};\end{aligned} \tag{136}$$

then

$$(137)\qquad \Big[(b^{(1)}+b^{(2)}+\cdots+b^{(s)})^n=\sum_{r=0}^{n}C_{n,r}\Big[b_{n-r}(b^{(1)}+\cdots+b^{(s-1)})^r,$$

and by (144)

$$\begin{aligned} &=\sum_{r=0}^{n}C_{n,r}\Big[b_{n-r}(b^{(1)}+\cdots+b^{(s-2)})^r\\ &\quad-\sum_{r=0}^{n}\frac{r}{s-2}C_{n,r}\Big[b_{n-r}(b^{(1)}+\cdots+b^{(s-2)})^r\\ &\quad-\sum_{r=0}^{n}rC_{n,r}\Big[b_{n-r}(b^{(1)}+\cdots+b^{(s-2)})^{n-1};\end{aligned}$$

or by previous methods this reduces to

$$(138)\qquad \begin{aligned} &\Big[(b'+\cdots+b^{(s)})^n\\ &=\Big[(b'+\cdots+b^{(s-1)})^n\\ &\quad-\frac{n}{s-2}\Big[(b^{(1)}+\cdots+b^{(s-2)})(b^{(1)}+\cdots+b^{(s-1)})^{n-1}\\ &\quad-\sum_{r=1}^{n}nC_{n-1,r-1}\Big[b_{n-r}(b'+\cdots+b^{(s-2)})^{n-1}\\ &=\Big[(b'+\cdots+b^{(s-1)})^n-\frac{n}{s-2}\cdot\frac{s-2}{s-1}(b^{(1)}+\cdots+b^{(s-1)})^n\\ &\quad-n\big[(b^{(1)}+\cdots+b^{(s-1)})^n,\end{aligned}$$

which is the result.

Employing (136), we find

$$(139)\qquad [(b+b'+b'')^n=\tfrac{1}{2}(n-1)(n-2)b_n+\tfrac{3}{2}n(n-2)b_{n-1}+n(n-1)b_{n-2}.$$

Repeated application of (135) gives easily

$$(140)\qquad \Big[(b+b'+\cdots+b^{(r)})^n=(-1)^r\frac{1}{r!}n(n-1)(n-2)\cdots(n-r)\sum_{i=0}^{r}\sigma_{ri}\frac{b_{n-i}}{n-i},$$

where the σ_{ri} are the elementary symmetric functions of the numbers $1, 2, 3, \cdots, r$, taken i at a time[20].

---

[20] This result is due to Lucas, Bulletin de la Société Mathématique de France, vol. 6 (1877), pp. 57–68. The proof given here is due to Dr. A. M. Mood. The proof of relation (144) was found independently by the writer.

Suppose $n \leqq r$ in (140); then all the terms will be zero except the one for $i=n$, hence

$$\Big[(b + b' + b^{(r)})^n = (-1)^r \frac{n!}{r!}(-1)^{r-n}(r-n)!\,\sigma_{rn},$$

hence

$$\sigma_{rn} = (-1)^n C_{r,n}(b + b' + \cdots + b^{(r)})^n. \tag{141}$$

This gives another form for (148) as follows:

$$\begin{aligned} &\Big[(b+b'+\cdots+b^{(r)})^n \\ &= (-1)^r \frac{1}{r!}\, n(n-1)\cdots(n-r)\sum_{i=0}^{r}(-1)^i C_{r,i}\Big[(b+b'+\cdots+b^{(r)})^i \frac{b_{n-i}}{n-i}. \end{aligned} \tag{142}$$

University of Texas,
Austin, Texas.

ON THE STRUCTURE OF DIFFERENTIAL POLYNOMIALS AND ON THEIR THEORY OF IDEALS

BY

HOWARD LEVI

In the first part of this paper a special class of differential ideals(1) is investigated. The results of this section are used in the following one to derive some structural properties of differential polynomials. The last part of the paper is devoted to a special differential ideal.

With the help of some conventions of notation, more precise indications of the scope of our work may be given. Let $\mathfrak{R}$ denote the ring of differential polynomials, with rational numbers for coefficients, in the unknown y. The special class of differential ideals studied in Part I is composed of those generated by y^p, where p is a positive integer. These ideals are among the most simple ideals encountered in the theory of differential equations. Viewed as algebraic entities, however, they are by no means trivial. We denote the ith derivative of y by y_i; $\mathfrak{R}$ thus appears as a polynomial ring with infinitely many indeterminates $y, y_1, y_2, \cdots$. Since the Hilbert basis theorem does not hold on $\mathfrak{R}$, one would expect almost any ideal in $\mathfrak{R}$ to be unruly. By introducing order relations into $\mathfrak{R}$ we have been able to proceed despite the absence of the basis theorem and to obtain fairly comprehensive results concerning these differential ideals. In particular a simple criterion for determining the membership in such an ideal of an element of $\mathfrak{R}$ is obtained which plays a fundamental role in Part II. This second part establishes the abstract counterparts of some results of J. F. Ritt concerning essential manifolds which figure in the decomposition of a manifold into irreducible ones. It has been found possible to present results which cover situations not discussed by him. The differential ideal discussed in Part III is that generated by uv, where u and v are unknowns. Among other properties, it is shown that this ideal has no representation as the intersection or product of two differential ideals, whose manifolds are respectively $u=0$ with v arbitrary, and $v=0$ with u arbitrary. This result owes its interest to the fact that the manifold of the equation $uv=0$ is evidently reducible into the union of the two manifolds just defined.

In a narrow sense, this paper is independent of other literature; the argu-

---

Presented to the Society February 22, 1941, under the title, *On the ideal theory and structure of differential polynomials*; received by the editors March 27, 1941.

(1) For terminology and bibliography, see *Semicentennial Addresses of the American Mathematical Society*, New York, 1938, pp. 35–58. The basic reference for the abstract theory of ideals of differential polynomials is H. W. Raudenbush, *Ideal theory and algebraic differential equations*, these Transactions, vol. 36 (1934), pp. 361–368.

ments make almost no appeal to outside sources. Less strictly, however, the writings of J. F. Ritt and H. W. Raudenbush, Jr., should be cited as furnishing both starting point and direction for this investigation. Indeed, the whole point of Part II lies in its connection with two papers of Ritt (more detailed reference is given in Part II). The congruence notation used in our work is that systematized by E. R. Kolchin[2]. Brackets [ ] and braces { } mean respectively the differential ideal and the perfect differential ideal generated by the set of elements they include. The congruences

$$a \equiv b \ [m, n, \cdots]$$
$$a \equiv b \ \{m, n, \cdots\}$$

mean, respectively, $a-b$ is in $[m, n, \cdots]$ and $a-b$ is in $\{m, n, \cdots\}$. We use the term "form" exclusively as an abbreviation for the term "differential polynomial."

PART I. THE DIFFERENTIAL IDEAL GENERATED BY y^p

THE FORM y^p AND ITS DERIVATIVES

1. Let p be any positive integer and let $A = y^p$. We investigate the differential ideal Σ generated by A. Denoting the ith derivative of A by A_i we see that Σ consists of all polynomials $E_0A + E_1A_1 + \cdots + E_rA_r$ where the E_i are any elements of $\mathfrak{R}$. It is sometimes convenient to let $y = y_0$, $A = A_0$.

We shall discuss power products in y and its derivatives, and make a few definitions for this purpose. Let $P = y^{p_0}y_1^{p_1} \cdots y_r^{p_r}$ be such a power product, the p_i being non-negative integers. The degree of P is defined as $\sum p_i$ and its weight is defined as $\sum ip_i$. A power product P is different from a power product $Q = y^{q_0}y_1^{q_1} \cdots y_s^{q_s}$ if some $p_i - q_i$ is different from zero. We understand that if $t > r$ then p_t, the exponent of y_t in P, is zero. If P is different from Q we say that P is *higher* than Q, and Q is *lower* than P, if the first nonzero difference $p_i - q_i$ is positive. If P is higher than Q and R is any power product, then RP is higher than RQ. If P is higher than Q and Q is higher than R, then P is higher than R. A power product P will be called an α *term*, if $p_i + p_{i+1} < p$; $i = 0, 1, 2, \cdots$. Every factor of an α term is an α term. Every power product not an α term will be called a β *term*. Every β term is divisible by an expression $y_i^r y_{i+1}^{p-r}$ with $r \leq p$.

2. The polynomial A_i, the ith derivative of A, is homogeneous of degree p and isobaric of weight i. A_i is a sum over j of terms $h_{ij}P_j$ where the h_{ij} are positive integers, and the P_j are power products of degree p and weight i. Each power product P_j of this weight and degree is present in A_i with a coefficient h_{ij} different from zero. In particular if $i = rp + s$ (r and s non-negative integers and $s < p$) the term $L_i = y_r^{p-s}y_{r+1}^s$ has the proper weight and degree

(2) *On the exponents of differential ideals*, Annals of Mathematics, (2), vol. 42 (1941), p. 741.

and is present effectively in A_i. It will be called the *leader* of A_i. We show that it is lower than any other term of A_i. It is certainly lower than any term involving a $y_k^{p_k}$ with $k<r$, $p_k>0$. Any term of A_i lower than L_i would thus be of the form $y_r^q y_{r+1}^{q_1} \cdots y_{r+t}^{q_t}$, with $q \leqq p-s$. This would imply that $q_1+q_2+\cdots+q_t \geqq s$. Thus the weight of such a term would be greater than $qr+(q_1+\cdots+q_t)(r+1)$ unless $q_2, \cdots, q_t$ were all zero. This last expression exceeds i if $q<p-s$. It follows that any term of A_i distinct from L_i must be higher than L_i.

Reduction of power products

3. We prove the following lemma.

Lemma 1.1. *For every β term F of $\mathfrak{R}$ there is a congruence*

$$F \equiv \sum_i h_i P_i \; [\Sigma]$$

where the P_i are α terms of the same weight and degree as F and the h_i are rational numbers (they may of course be zero).

F is divisible by the leader L_i of some A_i. Let c_i stand for the coefficient of L_i in A_i. Then $c_iL_i=A_i+(c_iL_i-A_i)$, where the terms in the parenthesis are higher than L_i, or are zero (if i is zero or unity). If $F=c_iL_iF'$, then

$$\begin{aligned} F &= F'A_i + F'(c_iL_i - A_i) \\ &\equiv F'(c_iL_i - A_i) \; [\Sigma]. \end{aligned} \tag{1.1}$$

All the terms of the right member of this congruence are higher than F and are of the same weight and degree as F. There may be some β terms among them. Each such term is likewise congruent to a sum of higher terms of the same weight and degree. In particular the lowest β term effectively present in (1.1) is congruent to such a sum. This term may be replaced in (1.1) by the appropriate combination of higher terms, yielding a new congruence for F free of this β term and all lower ones. Since there is only a finite number of power products of given weight and degree, this process eventually terminates; what remains in the right member is a linear combination of α terms with rational coefficients.

Canonical representations

4. The above lemma will be complemented by the fact established later that no linear combination of α terms with rational coefficients is in Σ unless the coefficients are all zero. In addition, a canonical representation for the elements of Σ will be obtained, in the following sense. Every element of Σ has a representation $E_0A+\cdots+E_sA_s$ but the same element may have different representations. A simple example of this is given by the polynomial $2y^2y_1$ which is in the ideal generated by y^2 and may be written $2y_1A$ or yA_1. Our canonical representation will be obtained by choosing the coefficients E_i

from a restricted set of polynomials, with the result that these coefficients are uniquely determined, while still furnishing representations for every element of Σ.

5. In securing the canonical representation for the elements of Σ we shall use forms H of the types

$$H = EA_{i_0}A_{i_1} \cdots A_{i_s}$$

where E is any power product in the y_i and the other factors of H constitute an arbitrary power product in A and its derivatives. It is convenient to write this latter power product as above, without using exponents, in such a way that $i_0 \leqq i_1 \leqq \cdots \leqq i_s$. H is homogeneous and isobaric. Its degree is the degree of E plus $(s+1)p$. Its weight is the weight of E plus $i_0+i_1+\cdots+i_s$. Evidently H is in Σ and conversely every element of Σ is a linear combination of such forms with constant coefficients. We order these forms in the following way: H is *higher* than $H'=E'A_{j_0}A_{j_1}\cdots A_{j_r}$ if either

(a) $A_{i_0}A_{i_1}\cdots A_{i_s}$ is higher than $A_{j_0}A_{j_1}\cdots A_{j_r}$ when both expressions are considered as power products in the A_i and are compared by the method used for power products in the y_i, or

(b) $r=s$, $i_k=j_k$ $(k=0, 1, \cdots, s)$ and E is higher than E' in the sense previously explained.

It should be emphasized that what we order are the symbols used to denote the forms rather than the forms themselves. For instance, for $A=y^2$, criterion (a) implies that $H=A\cdot A$ is higher than $H'=y^2A$ even though H and H' both denote the same form y^4. Thus an expression H is to be considered as different from H' for purposes of ordering, if the set $(i_0, i_1, \cdots, i_s)$ is different from the set $(j_0, j_1, \cdots, j_r)$ or if E is different from E'. We do not insist that the represented forms be different. On the other hand, *equations* connecting $H, H', \cdots$, are to refer in the usual way to the forms denoted by the symbols.

Evidently, of two different expressions H, H', one must be higher than the other. It is clear that our ordering is transitive. Furthermore, if H is higher than H' and G is any power product in the y_i, then GH is higher than GH'.

6. We now introduce the notion of a γ term. An expression $H=EA_{i_0}A_{i_1}\cdots A_{i_s}$ will be called a γ *term* if both (a) and (b) below hold.

(a) $i_s \geqq sp$,

(a_1) if $i_s>sp$ then E is an α term in the letters $y_{s+1}, y_{s+2}, \cdots$,

(a_2) if $i_s=sp$ then E is any power product in the letters $y_s, y_{s+1}, \cdots$.

(b) If $s>0$ then $i_k>kp$, $k=0, 1, \cdots, s-1$.

7. The role of these γ terms is revealed by the following lemma.

LEMMA 1.2. *Every expression H is equal to a sum $\sum r_iR_i$, where the R_i are γ terms of the same weight and degree as H. The r_i are rational numbers.*

This implies that every element of Σ is a linear combination of γ terms with rational coefficients. We shall see later that such a sum is zero only if all the coefficients are zero.

Our proof will consist mainly in showing that if H itself is not a γ term it is a linear combination with constant coefficients of γ terms and of expressions $H', H'', \cdots, H^{(i)}$; the expressions $H^{(i)}$ all being *higher* than H and of the same weight and degree as H. Once this is accomplished the proof can quickly be completed. By replacing the lowest $H^{(i)}$ by its linear combination of γ terms and expressions $H_j^{(i)}$ we obtain for H a new linear combination of γ terms and expressions $H_j^{(i)}$ which is free of that lowest $H^{(i)}$ and all lower ones. A finite number of repetitions of this procedure yields a linear combination of γ terms for H.

8. To devise methods for obtaining this sum of higher terms we consider the obvious equality $yA_1 = py_1A$ and those obtained by differentiating both members of this equation r times, $r = 1, 2, \cdots$. We obtain

$$yA_{r+1} + \sum_{i=1}^{r} C_{r,i}y_iA_{r+1-i} = p\sum_{i=1}^{r+1} C_{r,i-1}y_iA_{r+1-i}. \tag{1.2}$$

The symbols $C_{i,j}$ in equation (1.2) are binomial coefficients. This equation and the original equality together express yA_s with $c>0$ as a sum $\sum c_{si}y_iA_{s-i}$ where i runs through all positive integers not greater than s. An analogous expression may be obtained for y_kA_s, where k and s are positive integers and where $s > kp$. Let r and k be positive integers with the property $r+1-k>kp$. By subtracting the coefficient of y_kA_{r+1-k} in the right member of (1.2) from its coefficient in the left member we obtain

$$\frac{r!}{(k-1)!(r-k)!}\left(\frac{1}{k} - \frac{p}{r+1-k}\right).$$

This number is not zero, since $r+1-k>kp$. Thus under these circumstances y_kA_{r+1-k} is effectively present in (1.2). Equations (1.2) show, then, that y_kA_s with $s > kp$ may be written in the form

$$\sum_{i=1}^{s} d_{ki}y_{k+i}A_{s-i} + \sum_{i=1}^{k} e_{ki}y_{k-i}A_{s+i}$$

where the d_{ki} and e_{ki} are rational numbers which depend on s as well as on i and k. Observe that in the first of these sums the subscript of each A_{s-i} is less than s and in the second the subscript of each y_{k-i} is less than k.

9. We use these equations to derive a useful fact about expressions $y_kA_{i_0}A_{i_1}\cdots A_{i_s}$ when s and k are non-negative integers with $k \leqq s$ and $i_m > mp$, $m = 0, \cdots, s$. We show that such an expression is equal to a sum of certain products F_jG_j where the F_j are forms, and the G_j are power products in the A_i. It will be seen that the degree of each G_j in the A_i does not

exceed $s+1$, and that each G_j as a power product in the A_i is higher than $A_{i_0} \cdots A_{i_s}$. For $s=0$, $i_0>0$ we already know that yA_{i_0} is a sum of such products, namely $c_{i_0j}y_jA_{i_0-j}$. We establish the result for $s>0$ by induction. It follows from equations (1.2) that $y_kA_{i_0}A_{i_1} \cdots A_{i_s}$ is equal to

$$\left(\sum_i d_{ki}y_{k+i}A_{i_s-i}\right)A_{i_0} \cdots A_{i_{s-1}} + \left(\sum_i e_{ki}y_{k-i}A_{i_s+i}\right)A_{i_0} \cdots A_{i_{s-1}}.$$

The terms in the first group meet our requirements since each contains a power product in the A_i which is higher than $A_{i_0}A_{i_1} \cdots A_{i_s}$. Each term of the second group contains a factor $y_{k-i}A_{i_0} \cdots A_{i_{s-1}}$. Assume the result true for all integers less than s. Such a factor is then equal to a sum of products F_jG_j where the G_j are power products in the A_i which are higher than $A_{i_0} \cdots A_{i_{s-1}}$ and whose degrees in the A_i do not exceed s. Consequently the G_j are all higher than $A_{i_0}A_{i_1} \cdots A_{i_{s-1}}A_{i_s}$. By letting $F_j' = e_{ki}A_{i_s+i}F_j$ it is seen that the terms in the second group are likewise equal to a sum of the required type.

10. We are now in a position to carry out the proof. An expression H which is not a γ term must fail to satisfy at least one of the conditions (a) and (b). We enumerate the various possibilities and show how for each one the required sum of higher terms may be obtained. It is both permissible and convenient first to discuss those terms $H=EA_{i_0} \cdots A_{i_s}$ which satisfy (b) but not (a), and then to give a complete discussion for those expressions which do not satisfy (b). We follow this plan.

Suppose H satisfies (b), $i_s>sp$ and (a_1) is not satisfied. If E involves only the letters $y_{s+1}, y_{s+2}, \cdots$, it must be a β term. It must then be divisible by the leader L_f of some A_f. As in the proof of Lemma 1.1, we have $E=cE'A_f+F$ where c is a constant and F is a form every one of whose terms is higher than E. Consequently

$$H = cE'A_{i_0} \cdots A_{i_s}A_f + FA_{i_0} \cdots A_{i_s}.$$

The first term in the right member of this equation is the product with c of an expression H' which by criterion (a) above is higher than H. The rest of the right member of this equation consists of a linear combination with rational coefficients of terms $H^{(i)}$ all higher than H by criterion (b). We therefore have the required sum of higher terms.

If $i_s>sp$, if (a_1) is not satisfied and if E contains effectively some letter y_k with $k\leqq s$, we first write $E=y_kE'$. Because H is supposed to satisfy (b), we know that $H'=y_kA_{i_0} \cdots A_{i_s}$ is a sum of products F_jG_j as described earlier. It follows that H is a sum of products $(E'F_j)G_j$. Since the G_j are power products in the A_i whose degrees do not exceed $s+1$ and which are higher than $A_{i_0} \cdots A_{i_s}$, it follows from criterion (a) that all the expressions $H^{(i)}$ in each product $(E'F_j)G_j$ are higher than H. This disposes of expressions H which satisfy (b) but not (a_1). If an H satisfying (b) does not satisfy (a_2) and is such

that $i_s = sp$, then its coefficient E must contain effectively some letter y_k with $k < s$. Let $E = y_k E'$ and consider $y_k A_{i_0} \cdots A_{i_{s-1}}$. It equals a sum of products $F_j G_j$ where the G_j are power products in the A_i of degree not more than s and which are higher than $A_{i_0} \cdots A_{i_{s-1}}$. What is important for us, is that the G_j are consequently also higher than $A_{i_0} A_{i_1} \cdots A_{i_{s-1}} A_{i_s}$. It follows that $H = (A_{i_s} E') y_k A_{i_0} \cdots A_{i_{s-1}}$ is a sum of higher expressions of the required sort.

There remains the case of an H which does not satisfy (b). Let r be the smallest integer for which $i_r \leqq rp$ so that if $r > 0$ then $i_m > mp$, $m = 0, \cdots, r-1$. Our procedure depends on whether $i_r = rp$ or $i_r < rp$. If $i_r = rp$ let $EA_{i_{r+1}} \cdots A_{i_s}$ be expanded into a form $F = \sum h_i E_i$, the h_i being constants and the E_i power products. We have $H = (\sum h_i E_i) A_{i_0} \cdots A_{i_r}$. We consider the expressions $H_i = E_i A_{i_0} \cdots A_{i_r}$, noting that they all satisfy (b). Certain of the E_i may be free of the letters $y, y_1, \cdots, y_{r-1}$. For these E_i the corresponding H_i are γ terms and require no further discussion. If $r = 0$, all the E_i have this property and all the H_i are γ terms. On the other hand an E_i which contains effectively some y_k with $k < r$ leads to an H_i which is not a γ term. Such an H_i satisfies (b) but not (a_2). As we have seen, such an H_i is a sum of terms $q_j H_j^{(i)}$ where the q_j are constants and the expressions $H_j^{(i)}$ are all higher than H_i. This of course does not itself imply that the $H_j^{(i)}$ are higher than our original H. But by recalling that the $H_j^{(i)}$ must each contain a power product in the A_i which is higher than $A_{i_0} A_{i_1} \cdots A_{i_r}$ and whose degree does not exceed $r+1$, we see that the $H_j^{(i)}$ are actually higher than H by criterion (a). Our procedure for an H which does not satisfy (b) and for which $i_r < rp$ is the following. We note that r must be greater than zero, since i_r is non-negative. Let the form $EA_{i_r} A_{i_{r+1}} \cdots A_{i_s}$ be expanded as above into the form $F = \sum h_i E_i$. The fact that $i_r < rp$ implies that every term of A_{i_r}, and consequently every term of F, contains effectively some y_k with $k < r$. Then $H = (\sum h_i E_i) A_{i_0} \cdots A_{i_{r-1}}$ is a linear combination of expressions $H_i = E_i A_{i_0} \cdots A_{i_{r-1}}$ which satisfy (b) but do not satisfy the requirement of (a_1) which asks that E be free of $y, y_1, \cdots, y_r$. It is easy to see how the methods of the previous case apply here, and we omit the details of showing that H must be a sum of terms $h_j H_j^{(i)}$ where $H_j^{(i)}$ is higher than H.

11. We have now carried out our program of showing that each H not a γ term is a linear combination of expressions H_i, those H_i which are not γ terms being higher than H. The remarks made at the outset of the proof suffice to establish the lemma.

The fundamental lemma

12. We prove the following lemma.

Lemma 1.3. *Let d and w be positive integers. The number n_γ of γ terms of degree d and weight w does not exceed the number of n_β of β terms which have this weight and degree.*

It will be shown later that we actually have $n_\gamma = n_\beta$. We remind the reader that in computing n_γ one counts the number of distinct symbols which stand for γ terms without considering whether or not the symbols stand for distinct forms.

The proof will consist in associating a unique β term of degree d and weight w with each γ term of this degree and weight. The association will be such that to different γ terms there will correspond different β terms.

13. We require a few more definitions. Let $\mathfrak{R}_i$ denote the ring of polynomials with rational coefficients in the letters $y_i, y_{i+1}, \cdots, i=1, 2, 3, \cdots$, and let $\mathfrak{R}_0$ denote our original ring $\mathfrak{R}$. Let t be any non-negative integer. A form EA_i where $i \geqq tp$ and E is a power product in y and its derivatives will be called an expression K_t if the appropriate one of the following three conditions is satisfied by E.

(i) If $i=tp$, E is any power product of $\mathfrak{R}_t$.

(ii) If $tp<i\leqq(t+1)p$, E is any power product of $\mathfrak{R}_{t+1}$.

(iii) If $(t+1)p<i$, E is a special power product of $\mathfrak{R}_{t+1}$. Let $E=y_{t+1}^{a_1}\cdots y_{t+r}^{a_r}$. We ask that there exist an integer k for which $E'=y_{t+1}^{a_1}\cdots y_{t+k}^{a_k}$ is an α term, and in addition we require for this k that $i\leqq(t+k+1)p-(a_1+\cdots+a_k)$.

Under condition (iii) any α term of $\mathfrak{R}_{t+1}$ is acceptable as a coefficient E, for $(t+k+1)p$ increases with k, whereas for large k the exponent a_k is zero, so that $a_1+\cdots+a_k$ remains unchanged. On the other hand under condition (iii) an admissible E need not be an α term. Once a suitable k is found, no restriction whatever is made on the letters $y_{t+k+1}, y_{t+k+2}, \cdots$.

14. We now describe a process by which each such expression K_t determines a β term F of $\mathfrak{R}_t$. Let $K=EA_i$ be a definite expression K_t. It comes under one of (i), (ii), (iii).

If K comes under (i) we have $i_t=tp$. Let

$$F = y_t^p E.$$

Clearly F is a β term of $\mathfrak{R}_t$ having the same weight and degree as K.

If K comes under (ii) then $tp<i\leqq(t+1)p$. Let $b=(t+1)p-i$. Then b is a non-negative integer and $b<p$. Let

$$F = y_t^b y_{t+1}^{p-b} E.$$

F is a β term of $\mathfrak{R}_t$ obtained by replacing A_i in K by the term $y_t^b y_{t+1}^{p-b}$. The degree of this term is p and its weight is $(t+1)p-b=i$. Thus F has the same weight and degree as K. For this case E does not contain the letter y_t so that the exponent of y_t in F is b which is less than p. This distinguishes the F obtained from a K which comes under (ii) from that obtained from a K which comes under (i). However, for both cases the sum of the exponents of y_t and y_{t+1} in F is at least p.

If K comes under (iii) then $(t+1)p<i$. We define

$$s_0 = (t+1)p,$$
$$s_f = (t+f+1)p - (a_1 + \cdots + a_f), \qquad f = 1, \cdots, k.$$

We have $s_f - s_{f-1} = p - a_f$, $f = 1, \cdots, k$. By hypothesis $y_{t+1}^{a_1} \cdots y_{t+k}^{a_k}$ is an α term and in particular each a_f is less than p. Thus each $p-a_f$ is positive, so that $s_0 < s_1 < \cdots < s_k$. Since by hypothesis $i \leqq s_k$ there is an integer m which is such that $1 \leqq m \leqq k$ and for which $s_{m-1} < i \leqq s_m$. Let $b = s_m - i$. Then b is a non-negative integer. Since $b < s_m - s_{m-1}$ and $s_m - s_{m-1} = p - a_m$ we have $b + a_m < p$. Let

$$F = y_t^{a_1} \cdots y_{t+m-1}^{a_m}(y_{t+m}^{b} y_{t+m+1}^{p-b})G \tag{1.3}$$

where $G = y_{t+m+1}^{a_{m+1}} \cdots y_{t+r}^{a_r}$. F is a β term of $\mathfrak{R}_t$. In the transition from K to F the expression A_i is suppressed, $y_{t+j}^{a_j}$ is replaced by $y_{t+j-1}^{a_j}$, $j = 1, \cdots, m$, and the term $y_{t+m}^{b} y_{t+m-1}^{p-b}$ is introduced. G is carried over unchanged. The first operation lowers the weight by i and the degree by p. The second lowers the weight by $a_1 + \cdots + a_m$ and does not change the degree. The introduction of $y_{t+m}^{b} y_{t+m+1}^{p-b}$ augments the degree by p and the weight by $(t+m+1)p-b$. Since

$$(t + m + 1)p - b = s_m + (a_1 + \cdots + a_m) - b = i + (a_1 + \cdots + a_m) \tag{1.4}$$

we see that the net effect of these alterations is to leave the weight and degree unchanged. Note that F contains the factor $y_t^{a_1} \cdots y_{t+m-1}^{a_m} y_{t+m}^{b}$ and that its other letters all have subscripts which exceed $t+m$. This factor is an α term, since $b + a_m < p$ and $y_t^{a_1} \cdots y_{t+m-1}^{a_m}$ is an α term by hypothesis. Since m is positive, it follows that the sum of the exponents of y_t and y_{t+1} in F is less than p. This is a characteristic property of a term F obtained from an expression K which comes under (iii).

15. We have described a procedure for obtaining from any expression K_t a definite β term F of $\mathfrak{R}_t$. We shall investigate this procedure further in order to obtain two useful facts. The first is that by this process different β terms F are assigned to different expressions K. The second is that when $t>0$, then for any integer h such that $(t-1)p < h \leqq i$, the expression FA_h is an expression K_{t-1}. In other words if F is obtained from any K_t in the manner set forth above, then, if $t>0$ and h is as above, FA_h admits one of the three characterizations, (i), (ii), (iii), where the discussion is referred to the integer $t-1$ instead of t. In deriving the first property of the term F we need only show that a β term F cannot be obtained from two different expressions K_t which both come under the same condition of the three listed. This simplification is due to the fact that in describing the procedure it was pointed out how one could infer from a given F which of the three conditions governed the K which determined it. We now list the three possibilities for F and verify the two statements for each one.

16. Let F be determined by an expression $K=EA_i$ which comes under (i). Then $F=y_t^pE$, so that given F one can find E. Since for this case $i=tp$, there is only one EA_i which could lead to F. This proves the first statement. To prove the second, let h be any integer such that $(t-1)p<h\leqq i$. We can easily verify that FA_h is an expression K_{t-1}, coming under (ii). In the case at hand $i=tp$, so that we have $(t-1)p<h\leqq tp$. In addition F is a power product of $\mathfrak{R}_t$. These are precisely the requirements of condition (ii).

17. Now let F be determined by $K=EA_i$, K coming under (ii). We have $F=y_t^by_{t+1}^{p-b}E$ with $0\leqq b<p$ and with E free of $y, y_1, \cdots, y_t$. Again it is obvious that F determines E uniquely and that the subscript i of A_i can also be uniquely determined from the equation $(t+1)p-b=i$. Thus only one expression EA_i can yield F by our procedure. Suppose that $t>0$ and that h is some integer for which $(t-1)p<h\leqq i$. If $h\leqq tp$ then FA is an expression K_{t-1} coming under (ii), since F is in $\mathfrak{R}_t$. If $h>tp$ we show that FA_h is also an expression K_{t-1} but that it then comes under (iii). The inequality $h\leqq i=(t+1)p-b$ enables us to draw this conclusion. The α term required by (iii) is simply y_t^b; the integer k is unity.

18. The case in which $K=EA_i$ comes under (iii) remains. The F which it determines is displayed in (1.3). We noted above that F contains as a factor the α term $y_t^{a_1}\cdots y_{t+m-1}^{a_m}y_{t+m}^b$. In an obvious sense this factor is the "largest" α term which can be split off from F. More precisely, given an $F=y_t^{b_0}y_{t+1}^{b_1}\cdots y_{t+s}^{b_s}$ determined by an expression K_t which comes under (iii), if one chooses the largest g such that $y_t^{b_0}y_{t+1}^{b_1}\cdots y_{t+g}^{b_g}$ is an α term, this last power product will be identical with $y_t^{a_1}\cdots y_{t+m-1}^{a_m}y_{t+m}^b$. We recall that in passing from $K=EA_i$ to F we divided the letters of E into two classes; the letters of one class were replaced by others, and the letters in the other were carried over unaltered. What we have just shown is that given an F determined by a K which comes under (iii) it is possible to determine exactly which letters were in each of the classes. The weight of the A_i involved in K may be computed from

$$i = (t + m + 1)p - (a_1 + \cdots + a_m) - b.$$

Thus, given such an F it is possible to reconstruct unequivocally the expression K from which it was obtained. This establishes the first property for a K coming under (iii). Assuming now that $t>0$, we proceed to establish the second. Let h be an integer such that $(t-1)p<h\leqq i$. We show that FA_h is an expression K_{t-1}. If $h\leqq tp$, FA_h is clearly an expression K_{t-1} coming under (ii), since F is a power product of $\mathfrak{R}_t$. If $h>tp$, we show that FA_h is an expression K_{t-1} coming under (iii). To do this we must produce an α term and an integer k as described in (iii). Let $F=y_t^{b_0}\cdots y_{t+s}^{b_s}$. Since our calculations are now based on the integer $t-1$, the integer k is required to have the property $(t+k)p-(b_0+\cdots+b_{k-1})\geqq h$. Since $h\leqq i$, we see from (1.4) that the α term $y_t^{a_1}\cdots y_{t+m-1}^{a_m}y_{t+m}^b$ and the integer $m+1$ have the required properties.

19. The proof of the lemma may now be completed. Let $H=EA_{i_0}A_{i_1}\cdots A_{i_s}$ be any γ term of degree d and weight w. H satisfies conditions (a) and (b) defining a γ term. We now show how these conditions make it possible to use the work immediately preceding to carry out our program of assigning a β term to every γ term.

Consider the form $K^{(s)}=EA_{i_s}$. If $i_s=sp$ then $K^{(s)}$ is an expression K_s coming under (i). If $i_s>sp$ it is readily seen that $K^{(s)}$ is likewise an expression K_s, only in this case it comes under (iii). In fact condition (a_1) requires E to be an α term of $\mathfrak{R}_{s+1}$ and it was pointed out above that an integer k of the type required by (iii) can always be found under these circumstances. Thus by splitting off EA_{i_s} from a γ term H we always obtain an expression K_s. Let the weight of E be w_s and its degree be d_s. $K^{(s)}$ determines a β term of $\mathfrak{R}_s$ by the procedure described above. Let it be denoted by $E^{(s)}$. Its weight is w_s+i_s and its degree is d_s+p. If $s=0$ we associate this β term with H. It has the same weight and degree as H because it has the same weight and degree as $K^{(s)}$ and for this case $K^{(s)}=H$.

If $s>0$ consider the expression $K^{(s-1)}=E^{(s)}A_{i_{s-1}}$. It follows from the definition of γ term that $(s-1)p<i_{s-1}\leqq i_s$. This inequality permits us to conclude that $K^{(s-1)}$ is actually an expression K_{s-1}. $K^{(s-1)}$ determines a β term $E^{(s-1)}$ of $\mathfrak{R}_{s-1}$ having the same weight $w_s+i_s+i_{s-1}$ and the same degree d_s+2p as $K^{(s-1)}$. If $s=1$ we associate this β term with H. It clearly has the same weight and degree as H.

If $s>1$ we continue in this way. We obtain a sequence $K^{(s)}, K^{(s-1)}, \cdots, K^{(0)}$ and a sequence $E^{(s)}, E^{(s-1)}, \cdots, E^{(0)}$. The sequences are obtained from H by successive applications of the procedure described for obtaining β terms from expressions K_t. Each $K^{(f)}=E^{(f+1)}A_{i_f}$. Each $E^{(f)}$ is the β term of $\mathfrak{R}_f$ determined by $K^{(f)}$. The weight of both $K^{(f)}$ and $E^{(f)}$ is $w_s+i_s+\cdots+i_f$. The degree of both $K^{(f)}$ and $E^{(f)}$ is $d_s+(s-f+1)p$. The sequences are to be continued until $K^{(0)}$ and $E^{(0)}$ are reached. $E^{(0)}$ is a β term of $\mathfrak{R}=\mathfrak{R}_0$ having the same weight and degree as H. We associate $E^{(0)}$ with H.

20. We now prove that if $H_1=E_1A_{j_0}A_{j_1}\cdots A_{j_r}$ is a γ term different from H, then the β term $E_1^{(0)}$ assigned to it in this way must be different from $E^{(0)}$. H_1 determines the two sequences $K_1^{(r)}, K_1^{(r-1)}, \cdots, K_1^{(0)}$ and $E_1^{(r)}, E_1^{(r-1)}, \cdots, E_1^{(0)}$. We know that each $E^{(f)}$ is determined by at most one $K^{(f)}=E^{(f+1)}A_{i_f}$. We conclude that if $E^{(0)}=E_1^{(0)}$, then for every f for which the symbols are defined, $E^{(f)}=E_1^{(f)}$ and $K^{(f)}=K_1^{(f)}$. If $s=r$ we have immediately that $H=H_1$. Suppose $s\neq r$ and, say, $s<r$. We show that it is impossible to have $E^{(0)}=E_1^{(0)}$ under this assumption. This last equality implies that $K^{(s)}=K_1^{(s)}$, that is, that $EA_{i_s}=E_1^{(s+1)}A_{j_s}$. $E_1^{(s+1)}$ is determined by $K_1^{(s+1)}$ and is consequently a β term. From the definition of γ term E may only be a β term if $i_s=sp$. Thus $j_s=sp$. This is impossible since H_1 is a γ term and for such terms we have $j_f=fp$ only if $f=r$, whereas here we have $j_s=sp$ and $s<r$.

We have shown that every γ term determines a β term of $\mathfrak{R}$ having the

same weight and degree, and that distinct γ terms determine distinct β terms. This proves the lemma.

THE STRUCTURE OF THE IDEAL OF y^p

21. We can now prove the following lemma.

LEMMA 1.4. *Let d and w be positive integers. Let n_α denote the number of α terms P_i of degree d and weight w, let n_β denote the number of β terms Q_j of this degree and weight, and let n_γ denote the corresponding number of γ terms R_k. Then $n_\beta = n_\gamma$, and a relation*

$$\sum_{i=1}^{n_\alpha} p_i P_i + \sum_{k=1}^{n_\gamma} r_k R_k = 0$$

where the p_i and r_k are rational numbers implies that all the p_i and r_k are zero.

Let Q_j be any β term mentioned in the statement of the lemma. By Lemma 1.1 we have

$$Q_j \equiv \sum_i p_{ji} P_i \ [\Sigma], \qquad j = 1, \cdots, n_\beta.$$

Therefore by Lemma 1.2 we have

$$Q_j = \sum_k r_{jk} R_k + \sum_i p_{ji} P_i, \qquad j = 1, \cdots, n_\beta. \tag{1.5}$$

If n_γ were less than n_β some linear combination of the Q_j with rational coefficients would be a similar linear combination of the P_i. This is impossible, so that $n_\gamma \geqq n_\beta$. Applying Lemma 1.3 we see that $n_\gamma = n_\beta$.

22. Every R_k is by definition a homogeneous isobaric polynomial, so that we have

$$R_k = \sum_i a_{ki} P_i + \sum_j b_{kj} Q_j, \qquad k = 1, \cdots, n_\gamma. \tag{1.6}$$

Substituting the right member of (1.6) for R_k in (1.5), we obtain the identities $Q_j = Q_j$, $j = 1, \cdots, n_\beta$. Thus $|r_{jk}| \cdot |b_{kj}| = 1$ and it follows that both $|r_{jk}|$ and $|b_{kj}|$ are different from zero.

From (1.6) and the fact that $|b_{kj}| \neq 0$ we see that any linear combination of the R_k with constant coefficients not all zero must equal a similar linear combination of the P_i and Q_j which involves some Q_j effectively. If such a linear combination of terms R_k were expressible as a linear combination of terms P_i, we would have the absurd result that a linear combination of the Q_j with constant coefficients not all zero was a linear combination of the P_i.

23. THEOREM 1.1. *Let F be any element of $\mathfrak{R}$. F is expressible in the form*

$$F = \sum_i p_i P_i + \sum_k r_k R_k$$

where the P_i are α terms, the R_k are γ terms and the p_i, r_k are rational numbers(3). *For each F there is only one such expression.*

Let F be split up into a sum of homogeneous isobaric polynomials F_h in such a way that any two such polynomials have either different weights or different degrees. Then

$$F = \sum_h F_h$$

and the F_h are uniquely determined by F.

By means of equation (1.5) we have

$$F_h = \sum_i p_{hi}P_i + \sum_k r_{hk}R_k$$

where the d and w of Lemma 1.4 are the degree and weight of F_h. By adding the F_h we obtain

$$F = \sum_h \sum_i p_{hi}P_i + \sum_h \sum_k r_{hk}R_k,$$

and this is the desired equation.

24. To show that F does not have two distinct representations of this type we need only show that if

$$\sum_i p_iP_i + \sum_k r_kR_k = 0 \tag{1.7}$$

then the p_i and r_k are all zero.

Let some p_i or r_k be different from zero. The term which possesses such a coefficient must be cancelled in (1.7) by a sum of other terms of the same weight and degree. This contradiction to Lemma 1.4 establishes our result.

25. Corollary. *No linear combination of α terms with rational*(4) *coefficients is in Σ.*

Let d, w, n_α be as in the statement of Lemma 1.4. The quantity n depends on d and w.

Corollary. *The number of linearly independent* (mod Σ) *elements of $\mathfrak{R}$ which are homogeneous of degree d and isobaric of weight w is n_α.*

26. $\mathfrak{R}$ may be considered as an abelian group with operators, where the group "multiplication" is ordinary addition, and the operators are rational numbers. Theorem 1.1 implies that $\mathfrak{R}$ considered in this way is the direct sum of two groups. One of them is Σ. The other is the additive group generated by

(3) The same conclusion can be drawn if these symbols stand for any constants, or more generally if they are any elements of a domain of integrity which contains the rational numbers.

(4) See the note to Theorem 1.1.

the totality of α terms. A linearly independent basis for the first group is the totality of γ terms; the totality of α terms forms such a basis for the other.

27. We now determine circumstances under which the n_α of the corollary to Theorem 1.1 is zero. If for a given d and w this number is zero, then every homogeneous isobaric element of $\mathfrak{R}$ having this degree and weight is in Σ. In settling this question we consequently develop a method for establishing the membership in Σ of certain elements of $\mathfrak{R}$ based entirely on an examination of the weights and degrees of their constituent terms.

If d is less than p, every power product of degree d is an α term. To treat the case for which d is not less than p we write([8])

$$\mathfrak{S} = y^{p-1} y_2^{p-1} \cdots y_{2k}^{p-1} y_{2k+2}^{p-1} \cdots .$$

$\mathfrak{S}$ is a formal infinite product whose status in this discussion is that of a visual aid. Let d be a positive integer and write $d = a(p-1)+b$ (a and b non-negative integers with $0 < b \leqq p-1$). Let

$$S_d = y^{p-1} y_2^{p-1} \cdots y_{2a-2}^{p-1} y_{2a}^{b}.$$

S_d is an α term of degree d. It is obtained by taking the first d letters of $\mathfrak{S}$ and multiplying them together. We denote the weight of S_d by $w(p, d)$. We have

$$w(p, d) = a(a-1)(p-1) + 2ab.$$

$w(p, d)$ is defined for all integers p greater than unity, and for all positive integers d. Its values are always positive integers or zero. An easy calculation shows that $w(p, d)$ satisfies the difference equation $w(p, d)+2d = w(p, d+(p-1))$, and this fact is used in proving the following result.

28. THEOREM 1.2. *Let d, w and n_α be as in the statement of Lemma 1.4. A necessary and sufficient condition that $n_\alpha > 0$ is that $w \geqq w(p, d)$.*

In view of our earlier results this is equivalent to asserting that *every power product of degree d and weight $w < w(p, d)$ is in Σ and not every power product of degree d and weight $w \geqq w(p, d)$ is in Σ.*

29. The sufficiency proof is quickly disposed of. S_d is an α term of degree d and weight $w(p, d)$. Let

$$S_d^{(r)} = y^{p-1} y_2^{p-1} \cdots y_{2a-2}^{p-1} y_{2a}^{b-1} y_{2a+r}, \qquad r > 0.$$

([8]) In the remainder of Part I it is assumed that p exceeds unity. The two results enunciated there are seen to be trivially true for p equal to unity, if the weight function introduced at the end of §27 is defined to be plus infinity for p equal to unity and for all positive integral values of d.

$S_d^{(r)}$ is an α term of degree d and weight $w(p, d)+r$. Therefore we see that for any integer d and integer $w \geqq w(p, d)$ there are α terms of degree d and weight w.

30. We begin the necessity proof by observing that when d is less than p, $w(p, d)=0$. Consequently there are no power products of degree $d<p$ and weight $w<w(p, d)$. If our theorem were false there would be an integer $d \geqq p$, and an α term whose degree was d and whose weight was less than $w(p, d)$. We assume this to be the case and force a contradiction.

Let d ($\geqq p$) be the smallest integer for which there are α terms whose degree is d and whose weight is less than $w(p, d)$. Let P be an α term of degree d and weight w, where w is some integer such that $0 \leqq w<w(p, d)$. Let $P=EP'$, where E is that factor of P of degree $p-1$ which is higher than any other such factor. Then P' is of positive degree and is an α term. Furthermore, since P is an α term the definition of E insures that P' is free of y and y_1. Let the degree of P' be denoted by d' and its weight by w'. Clearly $w' \leqq w$ and $d'<d$. Let P'' be obtained from P' by replacing each letter y_i effectively present in P' by y_{i-2}. P'' is an α term whose degree is d' and whose weight is $w'-2d'$. Our assumption about the minimal character of d, when applied to P'' implies

$$w(p, d') \leqq w' - 2d'.$$

Using the difference equation satisfied by $w(p, d')$ this last inequality yields $w(p, d'+p-1) \leqq w'$. Since $d'+p-1=d$ and $w' \leqq w$ we now have $w(p, d) \leqq w$. This contradiction completes the proof.

The expression for a power product in the ideal of y^p

31. Having established the fact that certain power products are in Σ we may naturally inquire as to the number of derivatives of A needed to express them. This question may be precisely formulated in the following way. Let a power product P be in Σ. It is a linear combination of γ terms. Let the lowest of these be the γ term $EA_{i_0} \cdots A_{i_r}$. It is required to determine an upper bound for i_0. This question arises in the following section for a special class of power products. We settle it now for this special class.

Corollary. *Let r be a positive integer, let $d=(r+1)p-1$ and let w be a non-negative integer which does not exceed rd. Then every power product P of degree d and weight w is in Σ and is a linear combination, with forms for coefficients, of A and its derivatives of orders not exceeding rp.*

Let $d'=(r+1)(p-1)+1$. Certainly $d' \leqq d$. We first extract from P a factor P' of degree d' whose weight does not exceed rd'. This is made possible by the fact that the weight of P does not exceed rd. We then show that P', and hence P, is in Σ. Evaluation of $w(p, d')$ yields $r(r+1)(p-1)+2(r+1)$ which exceeds rd'. This shows that P' is in Σ. Actually for large r the weight of P is considerably smaller than $w(p, d)$. This additional restriction makes it pos-

sible to estimate relatively easily the number of derivatives of A required to express P.

The proof is by induction. When $r=1$, our assertion is that no more than p derivatives of A are required to obtain P. Any γ term of degree $2p-1$ is of the form A_iE. If i is zero no discussion is required. If i is greater than zero, E must be a power product of degree $p-1$ in the letters $y_1, y_2, \cdots$. The The weight of E is then at least $p-1$. If the weight of the γ term is not to exceed $2p-1$ it must be that $i\leqq p$. Assume now that the result is established for all integers less than some fixed integer r. Let $G=EA_{i_0}A_{i_1}\cdots A_{i_s}$ be a γ term of degree $d=(r+1)p-1$ and weight not greater than rd. It is to be shown that $i_0\leqq rp$. We need only consider the case in which $i_0>0$. For this case E must be free of y, since G is a γ term. Let E' be obtained from E by diminishing by unity the subscript of each y_t which appears in E. If $s>0$, consider $G'=E'A_{i_1-p}\cdots A_{i_s-p}$. G' is evidently a γ term of degree $d'=d-p$ and weight $w'=w-i_0-d+p$. If i_0 exceeded rp we should then have, using $w\leqq rd$, $w'<rd-rp-d+p$ or $w'<(r-1)d'$. Our induction hypothesis then applies to G' and shows that $i_1-p\leqq(r-1)p$ whence $i_1\leqq rp$. Since $i_0\leqq i_1$, the assumption $i_0>rp$ leads to a contradiction. If in G the integer s is zero, so that $G=EA_{i_0}$, a different procedure is required. E must be an α term in the letters $y_1, y_2, \cdots$, of degree $d-p$ and weight $w-i_0$. Consequently E' is an α term of degree $d'=d-p$ and weight $w'=w-i_0-d+p$. Again assume that $i_0>rp$. Using $w\leqq rd$ and $i_0>rp$ we have $w'<(r-1)d'$. Since d' is $rp-1$ it follows as in the outset of this proof that the weight of E' is too small for E' to be an α term. The hypothesis $i_0>rp$ must then be discarded and the induction is carried out.

Part II. Some theorems on the structure of differential polynomials

The low power theorem

32. Let $\mathfrak{J}$ be any differential domain of integrity which contains the rational numbers. Throughout this section when we refer to a form in the unknowns $u, v, \cdots, w$, we shall mean a differential polynomial in $u, v, \cdots, w$ whose coefficients are in $\mathfrak{J}$. Indeed the coefficients actually used are for the most part rational numbers. Our work involves auxiliary unknowns which may be specialized with great freedom, and it is with such specialization in mind that the above remarks are made.

33. Theorem 2.1. *Let*

$$F=\lambda y^p-\sum_{i=1}^{k}u_iB_i \tag{2.1}$$

be a form in $y, \lambda, u_1, \cdots, u_k$ where p is a positive integer and the B_i are power products in y and its derivatives of degree $p+1$. Then there is a positive integer s and a form

$$D = \lambda^s + H$$

where every term of H contains y or one of its derivatives effectively and where D is homogeneous of degree s in λ, $u_1, \cdots, u_k$ and their derivatives, such that for some positive integer d,

$$y^d D \equiv 0 \ [F].$$

The questions as to how large d and s need be, and how many derivatives of F are required to obtain y^dD are not answered precisely, but in the proof explicit upper bounds are given for each of these numbers.

34. This theorem is the abstract counterpart of certain results obtained by Ritt. It might be appropriate to discuss this relation before taking up the proof. If y, λ and the u_i are replaced by forms Y, L, U_i in the unknowns $v_1, \cdots, v_n$, then F and D go over into forms F' and D' in the v_i and $Y^dD' \equiv 0$ $[F']$. Let us take Y to be an algebraically irreducible form whose order in v_n is h, L to be a nonzero form not divisible by Y whose order in v_n does not exceed h, and the U_i to be any forms. Then, when $\mathfrak{J}$ is a ring of analytic functions, the relation $Y^dD' \equiv 0$ $[F']$ shows that the general solution of Y is an essential irreducible manifold in the manifold of F'[6]. A somewhat different result is obtained by specializing y as v_1, λ as a form $1+L$ where L vanishes for $v_i = 0$, $i = 1, \cdots, n$, and the u_i as any forms in the v_j. $\mathfrak{J}$ is again a ring of analytic functions. The relation $v_1^dD' \equiv 0$ $[F']$ shows that the solution $v_i = 0$, $i = 1, \cdots, n$, of F' is not contained in any irreducible manifold held by F' but not by v_1[7].

35. We now take up the proof. Let r be the maximum of the weights of the B_i. If r is zero or unity, each B_i is divisible by y^p, and F itself may be factored into a product $y^p(\lambda + H)$ of the required type. Assuming now that $r > 2$, let d be an integer such that the set of all power products in y and its derivatives of degree d whose weight does not exceed $(r-1)d$ is in the differential ideal generated by $A = y^p$. The work of the preceding section proves that there are such integers d. Let these power products be denoted by $P_1, \cdots, P_m$, and let the weight of P_f be denoted by w_f, $f = 1, \cdots, m$. Suppose all the P_f may be expressed as linear combinations of A and its derivatives of order not exceeding t. Then

$$P_f = \sum_{j=0}^{t} C_{fj} A_j, \qquad f = 1, 2, \cdots, m,$$

where the C_{fj} are forms in y. They are homogeneous of degree $d-p$ and isobaric of weight $w_f - j$. Using the fact that $\lambda^{q+1}A_q$ is in the differential ideal

(6) J. F. Ritt, *On the singular solutions of algebraic differential equations*, Annals of Mathematics, (2), vol. 37 (1936), pp. 555–560.

(7) J. F. Ritt, *On certain points in the theory of algebraic differential equations*, American Journal of Mathematics, vol. 60 (1938), p. 9. This paper will be referred to as OCP.

generated by λA and that it is a linear combination of λA and its first q derivatives([8]), we have for each P_f

$$\lambda^{t+1}P_f = \sum_{j=0}^{t}\sum_{g=0}^{j} C_{fj}L_g(\lambda A)_{j-g}.$$

Here the symbol $(\lambda A)_{j-g}$ means the $(j-g)$th derivative of λA, and L_g is a homogeneous form in λ of degree t. Referring to (2.1) we have for each P_f

$$\lambda^{t+1}P_f \equiv \sum_{j=0}^{t}\sum_{g=0}^{j} C_{fj}L_g\left(\sum_{i=1}^{k} u_iB_i\right)_{j-g} [F]$$

where $(\sum u_iB_i)_{j-g}$ means the $(j-g)$th derivative of the form inside the parentheses. The right member of this congruence is a linear combination of forms $T = C_{fj}(B_i)_h$, $h \leqq j$, whose coefficients are homogeneous forms in λ and the u_i. These forms T are homogeneous of degree $d+1$, since the B_i are all of degree $p+1$. The weight of each $(B_i)_h$ does not exceed $r+h$, and therefore the weight of each T does not exceed $w_f - j + r + h$. Using $w_f \leqq (r-1)d$ and $h \leqq j$, it follows that the weight of each T does not exceed $(r-1)d+r$. The forms T are thus linear combinations of power products y_cP_f where again the P_f are power products in the y_i of degree d and weight not exceeding $(r-1)d$. We need merely choose $c \geqq r$ if such a y_c appears in a term of the T, while if no such y_c is present effectively, then for any choice of y_c the statement is true.

36. What we have shown is that

$$\lambda^{t+1}P_f \equiv \sum_{j=1}^{m} E_{fj}P_j\ [F], \qquad f = 1, \cdots, m,$$

where the E_{fj} are forms in y, and the u_i. E_{fj} is homogeneous of degree t in λ, homogeneous of degree unity in the u_i and homogeneous of degree unity in the y_i. Transposing, we have a system of m linear congruences for the P_f. In the ith congruence the coefficients of P_j with $j \neq i$ is $-E_{ij}$ while the coefficient of P_i is $\lambda^t - E_{ii}$. It follows that

$$P_fD \equiv 0\ [F], \qquad f = 1, 2, \cdots, m,$$

where D is the determinant of the system of congruences. Clearly D is of the form $\lambda^{m(t+1)} + H$ where H vanishes for $y=0$. Since y^d is one of the P_f we have our result. Observe that H is homogeneous of degree $m(t+1)$ in λ and the u_i.

37. As a supplement to the proof the size of t and d will be examined. Reference to the work of Part I reveals that if $d = rp-1$ then t may be taken as $(r-1)p$. If, on the other hand, $d = r(p-1)+1$, the proof may also be car-

([8]) This is obvious for $q=0$. An easy induction establishes it for all positive integral values of q, for the product of λ^q with the qth derivative of A is a form $\lambda^{q+1}A_q$ plus terms all of which contain a factor $\lambda^q A_i$ with $i<q$.

ried out with possibly a larger value of t. Note that this value of d is in general smaller than the previous one.

GENERALIZATION FOR ONE UNKNOWN

38. This theorem may be extended in the following way. Let

$$F = \lambda y^{p_0} y_1^{p_1} \cdots y_r^{p_r} + \sum u_i B_i \tag{2.2}$$

be a form in y, λ and a finite number of unknowns u_i where $p_0, p_1, \cdots, p_r$ are non-negative integers with $p_r \neq 0$, and where the B_i are power products in the y_i. If the degree of each B_i in the letters $y_r, y_{r+1}, \cdots$, exceeds p_r, then Theorem 2.1 may be applied to F. It yields a relation

$$y_r^d((\lambda y^{p_0} y_1^{p_1} \cdots y_{r-1}^{p_{r-1}})^s + H) \equiv 0 \; [F],$$

where H vanishes for $y_{r+i} = 0$, $i = 0, 1, 2, \cdots$. It is obvious that F admits the solution $y = 0$. This relation shows that every irreducible manifold held by F which contains $y = 0$ and is not held by y_r must be held by the form in the outer parentheses. We are going to show how additional hypotheses on the B_i make it possible to draw a stronger conclusion.

THEOREM 2.2. *Let the F be given by* (2.2) *and let the B_i be such that for each integer f, $f = 0, 1, \cdots, r$, the degree of each B_i in the letters $y_f, y_{f+1}, y_{f+2}, \cdots$ exceeds $p_f + p_{f+1} + \cdots + p_r$. Then there is a positive integer s and a form*

$$D = \lambda^s + H$$

where every term of H contains y or one of its derivatives effectively and where D is homogeneous of degree s in λ and the u_i and their derivatives, such that for some positive integer d

$$y_r^d D \equiv 0 \; [F].$$

The stronger conclusion drawn here is that if λ is specialized as any non-zero element of $\mathfrak{J}$, then the solution $y = 0$ of F is contained in no irreducible manifold held by F but not by y_r.

39. The proof is by induction on r. For $r = 0$ this result is practically identical with Theorem 2.1. The only difference is that the B_i may be of degree greater than $p_0 + 1$ in the y_i. However, it is easy to see that by incorporating superfluous factors of the B_i into their coefficients we obtain a form for which Theorem 2.1 may be invoked. We assume the theorem established for integers from zero to $r-1$ inclusive and prove it for r.

40. We introduce new letters λ', u_i' and a new form F' in y_1, λ', u_i' in the following way. Let Q_i be that factor of B_i of degree p_0 which is higher than any other such factor (if p_0 is zero then Q_i is unity). Let B_i' be defined by $B_i = B_i' Q_i$. Let

$$F' = \lambda' y_1^{p_1} y_2^{p_2} \cdots y_r^{p_r} + \sum u_i' B_i'.$$

F' goes over into F when λ' is replaced in F' by λy^{p_0} and u_i' by $u_i Q_i$. It will be shown that F' satisfies the induction hypothesis for $r-1$. To do this it suffices to show that for all integers f, $f=1, 2, \cdots, r$, the degree of each B_i' in $y_f, y_{f+1}, \cdots$ exceeds $p_f + p_{f+1} + \cdots + p_r$. Suppose there were a B_i' and an f for which this were not so. Because of the way B_i' was defined, our hypothesis would then require that Q_i be of positive degree in some y_t with $t \geqq f$. Since Q_i is the highest factor of B_i of degree p_0, it follows that B_i' is free of $y, y_1, \cdots, y_{f-1}$. Thus the degree of B_i is simply p_0 plus the degree of B_i' in $y_f, y_{f+1}, \cdots$. Clearly the statement that the degree of B_i' in $y_f, y_{f+1}, \cdots$ does not exceed $p_f + p_{f+1} + \cdots + p_r$ conflicts with our assumption that the degree of B_i exceeds $p_0 + p_1 + \cdots + p_r$.

41. Under our induction hypothesis there is a form $D' = \lambda'^t + H'$ and an integer d_1 such that

$$y_r^{d_1} D' \equiv 0 \ [F']$$

where H' is homogeneous in λ' and the u_i' of degree t and vanishes for $y_1 = 0$. Let D' become D_1 when λ' and the u_i' are replaced as above. Then $y_r^{d_1} D_1$ is in $[F]$. D_1 is of the form $\lambda^t y^{p_0 t} + H_1$ where H_1 is homogeneous in λ and the u_i of degree t. Every term of H is of degree at least $tp_0 + 1$ in the y_i.

It follows from Theorem 2.1 that there is a form $D_2 = \lambda^w + H_2$, H_2 homogeneous in λ and the u_i of degree w and vanishing for $y = 0$, and an integer k such that $y^k D_2 \equiv 0 \ [D_1]$. By the result stated in the footnote to Theorem 1.1 we find that there are integers h and g such that $y_r^h D_2^g \equiv 0 \ [D_1]$. This same result used again shows that for sufficiently large d, $y_r^d D_2^g$ is in $[y_r^{d_1} D_1]$. This completes the proof, since $y_r^{d_1} D_1$ is in $[F]$, and the form $D = D_2^g$ and the integer d have the required properties.

A special case

THEOREM 2.3. *Let p be a positive integer, let $A = y^p$, and let A_i be the ith derivative of A. Let*

$$F = A - \sum_{i,j} C_{ij} y_i A_j$$

where i and j have some definite range and where the C_{ij} are forms in y. Then there is a form $D = 1 + H$, where H is a form in y which vanishes for $y = 0$, such that

$$y^p D \equiv 0 \ [F].$$

The distinctive feature of this result is that the integer p is available as the integer d of our previous work. For $p = 1$ this theorem is identical with a result obtained in a paper by Ritt and Kolchin([9]).

([9]) J. F. Ritt and E. R. Kolchin, *On certain ideals of differential polynomials*, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 895–898.

The device of continued substitution used in the first part of our proof is borrowed from this paper.

42. By hypothesis

$$A \equiv \sum_{i,j} C_{ij} y_i A_j \ [F]. \tag{2.3}$$

If in the right member of (2.3) A_j is replaced by the jth derivative of the whole right member, the result is a congruence

$$A \equiv \sum_{i,j,k} C_{ijk} y_i y_j A_k \ [F]. \tag{2.4}$$

For each C_{ij} of (2.3) which is different from zero, let the sum $i+j$ be computed. Suppose r is the largest of these sums. Then no sum $i+j+k$ for which there is a nonzero C_{ijk} exceeds $2r$. If in the right member of (2.4) the kth derivative of the whole right member of (2.3) is substituted for A_k, a new congruence for A is obtained. After $s-1$ iterations of this substitution, a congruence

$$A \equiv \sum C_{i_1 i_2 \cdots i_{s+1}} y_{i_1} y_{i_2} \cdots y_{i_s} A_{i_{s+1}} \ [F]$$

is obtained, where no sum $i_1+i_2+ \cdots +i_s$ with nonzero $C_{i_1 \cdots i_{s+1}}$ exceeds sr. If $s=(r+1)(p-1)+1$ then by our frequently used criterion, $y_{i_1} y_{i_2} \cdots y_{i_s} \equiv 0[A]$. Thus

$$A \equiv \sum D_{ij} A_i A_j \ [F] \tag{2.5}$$

where D_{ij} is a form in y.

43. We now compute the sum $i+j$ for each D_{ij} effectively present in (2.5). Let s be the largest of these sums. If (2.5) is differentiated s times, we obtain $s+1$ congruences, expressing A_k, $k=0, 1, \cdots, s$, as linear combinations of products A_iA_j whose coefficients are forms in y. Each product so obtained must contain an A_i with $i \leqq s$, for the differentiation introduced an increment of at most s to the sums $i+j$ and they did not exceed s at the outset. We have shown, then, that

$$A_i \equiv \sum_{j=0}^{s} K_{ij} A_j \ [F], \qquad i = 0, 1, \cdots, s,$$

where K_{ij} vanishes for $y=0$. Transposing, we have a system of $s+1$ linear homogeneous congruences for the A_i. In the ith congruence the coefficient of A_j with $j \neq i$ is $-K_{ij}$ while the coefficient of A_i is $(1-K_{ii})$. D, the determinant of the system, is of the form $1+H$ where H vanishes for $y=0$. Since $A_iD \equiv 0 \cdot [F]$; $i=0, 1, \cdots, s$ we have our result.

Generalizations for several unknowns

44. Our next result concerns systems of forms in the unknowns $y_1, \cdots, y_n, \lambda$ and a finite number of unknowns u_{ij}. As is customary, we shall denote the jth

derivative of y_i with the symbol y_{ij} and the ith derivative of λ by λ_i. The second subscript of the u_{ij} will not mean differentiation, but will simply indicate how these unknowns are displayed in rows and columns.

45. Let s be any positive integer not greater than n. We consider subsets $(i_1, i_2, \cdots, i_s)$ of $(1, 2, \cdots, n)$ where in each subset the numbers i_k are all different. If the binomial coefficient $C_{n,s}$ is denoted by q, there are exactly q such subsets. We suppose a number $j, j=1, \cdots, q$, assigned in any univocal manner to each such subset. We shall consider the system

$$F_j = \lambda y_{i_1}^{p_{j1}} y_{i_2}^{p_{j2}} \cdots y_{i_s}^{p_{js}} + \sum_i u_{ji} B_{ji}, \qquad j = 1, \cdots, q, \tag{2.6}$$

where the p_{ji} are positive integers, and the B_{ji} power products in $y_1, \cdots, y_n$ and their derivatives. We call $\lambda y_{i_1}^{p_{j1}} y_{i_2}^{p_{j2}} \cdots y_{i_s}^{p_{js}}$ the *first term* of F_j. For each form F_j we make the following assumptions concerning the degree of the B_{ji} in the unknowns $y_1, \cdots, y_n$ and their derivatives. These assumptions describe a relation whereby the B_{ji} dominate the first terms of the F_j and it should be understood that each B_{ji} is qualified in this way only by the first term of that form F_j which contains it. Let $(i_a, i_b, \cdots, i_f)$ be any (proper or improper) subset of $(i_1, i_2, \cdots, i_s)$. It is required of B_{ji} that

(1) either

(1a) B_{ji} be divisible by $y_{i_a}^{p_{ja}} y_{i_b}^{p_{jb}} \cdots y_{i_f}^{p_{jf}}$

or

(1b) the degree of B_{ji} in the unknowns $y_{i_a}, y_{i_b}, \cdots, y_{i_f}$, in the unknowns y_k not in the first term of F_j, and in the derivatives of all these unknowns exceeds $p_{ja}+p_{jb}+\cdots+p_{jf}$;

(2) the total degree of each B_{ji} must exceed

$$p_{j1} + p_{j2} + \cdots + p_{js}.$$

THEOREM 2.4. *Let Σ be the differential ideal generated by the F_j. Then there exists a form $D=\lambda^c+H$, where H is a form in the y_i, λ and the u_{ij}, which vanishes for $y_i=0$, $i=1, \cdots, n$, and an integer t such that for every form $V_j=y_{i_1}y_{i_2}\cdots y_{i_s}$*

$$V_j^t D \equiv 0 \ [\Sigma]. \tag{2.7}$$

H is homogeneous of degree c in λ and the u_{ji}.

46. Here, too, our result has considerable contact with Ritt's work in differential equations. Before taking up the proof a few remarks might be made concerning the content of this theorem from the standpoint of differential equations. The unknowns λ and the u_{ij} have been introduced as auxiliaries to facilitate the proof. For purposes of illustration we may suppose λ replaced by unity and the u_{ji} by any forms in the y_i. Equations (2.7) then have the appearance

$$V_j^t(1+H) \equiv 0\ [\Sigma], \qquad j = 1, \cdots, q.$$

Obviously Σ admits the solution $y_i=0$, $i=1, \cdots, n$. These equations show that any irreducible manifold held by Σ which contains $y_i=0$, $i=1, \cdots, n$, must be held by the system V_j, $j=1, \cdots, q$. Each irreducible manifold in this latter system is found by letting some $n-s+1$ of the unknowns y_i be zero, the remaining $s-1$ unknowns being arbitrary. In one extreme case, with $s=1$, the manifold of the system V_j is precisely $y_i=0$, $i=1, \cdots, n$. The essentiality of this solution in Σ was shown by Ritt(10). He also treated the case $s=n$, obtaining(11) the above conclusion as a consequence of an approximation theorem. The intermediate cases, that is, those in which $1<s<n$, appear here as new results, both from the abstract viewpoint and that of differential equations. The extreme cases owe their novelty to the fact that their proof is abstract. For the extreme case $s=n$, we reverse the procedure followed by Ritt. This result will be established first and then the analogue of Ritt's approximation theorem will be shown to follow from it. For $s=n$ and $\lambda=1$ we have a single form

$$F = y_1^{p_1} y_2^{p_2} \cdots y_n^{p_n} + \sum u_i B_i.$$

Our hypothesis now reduces to the statements that for each k each B_i is either divisible by $y_k^{p_k}$ or its degree in the y_{kj} exceeds p_k. The total degree of each B_i exceeds $p_1 + \cdots + p_n$. The conclusion is that there is a congruence

$$(y_1 \cdots y_n)^t(1+H) \equiv 0\ [F]$$

where H vanishes for $y_i=0$, $i=1, \cdots, n$. This special case of Theorem 2.4 will be used in our later work.

47. We prove Theorem 2.4 by induction on s, beginning with $s=1$ and n arbitrary. For this case our hypothesis states that Σ contains the forms

$$F_j = \lambda y_j^{p_j} + \sum_i u_{ji} B_{ji}, \qquad = 1, \cdots, n,$$

the B_{ji} being power products in the y_i and their derivatives whose total degree exceeds p_j.

Let r be the maximum of the weights of the B_{ji}. We understand that the weight of y_{jk} is k. Let p be the maximum of the p_j. Let $d=n(r+1)(p-1)+1$. Then every power product P_h in the y_{ij} of degree d whose weight w_h does not exceed $r((r+1)(p-1)+1)$ is in some $[y_i^{p_i}]$, for P_h is of degree at least $(r+1)(p-1)+1$ in at least one y_j, and then our earlier result applies. We now follow the procedure used in the proof of Theorem 2.1. We first multiply each P_h by λ^g so that the product is in some $[\lambda y_i^{p_i}]$ and after substitutions and re-

(10) OCP, pp. 5–7.
(11) OCP, p. 14.

arrangements similar to those used in the proof of that theorem we obtain the congruences

$$\lambda^g P_h \equiv \sum_{i=1}^{m} E_{hi} P_i \ [\Sigma], \qquad h = 1, 2, \cdots, m,$$

m being the number of P_h. The only difference is that now the E_{hi} are forms in $y_1, \cdots, y_n$, λ and the u_{ji} which need not be homogeneous in the y_i. It is still true that every term of each E_{hi} involves some y_k effectively. E_{hi} is homogeneous of degree $g-1$ in λ and is homogeneous of degree unity in the u_{ij}. After transposing, we see that

$$P_h D \equiv 0 \ [\Sigma], \qquad h = 1, \cdots, m,$$

where D, the determinant of the transposed system, is of the type described in the statement of this theorem. Since $y_1^d, y_2^d, \cdots, y_n^d$ are all to be found among the P_h we have our result.

48. Continuing with the proof we suppose that the theorem holds for all values of s from unity to some fixed integer, and proceed to show that it holds for the next integer. We shall denote this latter integer by s. In our induction assumptions the only restriction on n is that it be sufficiently large to insure that the statement of the theorem makes sense, that is, n is never less than s. It is otherwise arbitrary.

Referring to (2.6) select all those forms F_j whose first terms contain the letter y_1 effectively. Let p be the maximum of the exponents of y_1 in the first terms of these F_j. By multiplying certain of these F_j by a suitable power of y_1 we obtain a system of forms G_j each of whose first terms contains the letter y_1 exactly to the pth power. The set of forms G_j, being composed of some of these F_j and multiples by a power of y_1 of the others, is certainly in Σ. Furthermore the terms of the G_j which are not first terms satisfy the conditions of our hypothesis relative to the first term of the form which contains them. We now introduce new letters λ', u'_{ji} and new forms G'_j in these letters and $y_2, y_3, \cdots, y_n$. The G'_j will be defined so that their first terms will contain only $s-1$ letters. They will fulfill the conditions of our hypothesis, and will go over into the G_j when appropriate replacements are made for λ' and the u'_{ji}. We begin with

$$G_1 = \lambda y_1^p y_2^{p_{12}} \cdots y_s^{p_{1s}} + \sum u_{1i} \bar{B}_{1i}.$$

The $\bar{B}_{1i}$ are either the original B_{1i} or multiples of them by a power of y_1. As a first step we divide the $\bar{B}_{1i}$ into two classes, those which are divisible by $y_2^{p_{12}} \cdots y_s^{p_{1s}}$ and those which are not. A $\bar{B}_{1i}$ which is divisible by this term may be written as $L_{1i}\, y_2^{p_{12}} \cdots y_s^{p_{1s}}$. Our hypothesis reveals that the degree of L_{1i} in $y_1, \cdots, y_n$ and their derivatives exceeds p. We now have

$$G_1 = \left(\lambda y_1^p + \sum_k u_{1k} L_{1k}\right) y_2^{p_{12}} \cdots y_s^{p_{1s}} + \sum u_{1i} \bar{B}_{1i}$$

where the second summation is performed over all those $\bar{B}_{1i}$ which are in the second class. We further subdivide the $\bar{B}_{1i}$ of the second class into those which contain a factor in y_1 and its derivatives of degree p and those which do not. For a $\bar{B}_{1i}$ of the first kind let Q_{1i} be any such factor and let $\bar{B}_{1i}=Q_{1i}B_{1i}'$. For $\bar{B}_{1i}$ of the second kind the total multiplicity of y_1 and its derivatives in $\bar{B}_{1i}$ is some number $q_{1i}<p$. Let all these letters be split off from $\bar{B}_{1i}$ and multiplied by any factor of $\bar{B}_{1i}$ which contains only the unknowns $y_{s+1}, \cdots, y_n$ and whose degree in these unknowns and their derivatives is $p-q_{1i}$. Our hypothesis permits us to construct such a product for each $\bar{B}_{1i}$ of the second kind. Denoting this product by Q_{1i} we have in Q_{1i} a power product in $y_1, y_{s+1}, \cdots, y_n$ and their derivatives of degree p. Here too we let $\bar{B}_{1i}=Q_{1i}B_{1i}'$. Let

$$G_1' = \lambda_1 y_2^{p_{12}} \cdots y_s^{p_{1s}} + \sum u_{1i}' B_{1i}'.$$

When λ_1 is replaced by $\lambda y_1+\sum u_{1k}L_{1k}$ and u_{1i}' by $u_{1i}Q_{1i}$ then G_1' goes over into G_1.

49. It is readily seen that if $(a, b, \cdots, f)$ is any subset of $(2, 3, \cdots, s)$ then for each B_{1i}' either

(a) B_{1i}' is divisible by $y_a^{p_{1a}} y_b^{p_{1b}} \cdots y_f^{p_{1f}}$

or

(b) the degree of B_{1i}' in $y_a, y_b, \cdots, y_f, y_{s+1}, \cdots, y_n$ and their derivatives exceeds $p_{1a}+p_{1b}+\cdots+p_{1f}$.

For a B_{1i}' obtained from a $\bar{B}_{1i}$ of the first kind this is part of our initial hypothesis, since in this case B_{1i}' is the same as $\bar{B}_{1i}$ as far as the unknowns $y_2, y_3, \cdots, y_n$ are concerned. As for a B_{1i}' obtained from a $\bar{B}_{1i}$ of the second kind, recall that $\bar{B}_{1i}=Q_{1i}B_{1i}'$ where Q_{1i} is of degree p in $y_1, y_{s+1}, \cdots, y_n$ and their derivatives and where B_{1i}' is free of y_1 and its derivatives. Since $\bar{B}_{1i}$ is not divisible by $y_1^p y_a^{p_{1a}} y_b^{p_{1b}} \cdots y_f^{p_{1f}}$ (because its degree in y_1 is less than p) our original hypothesis requires that the degree of $\bar{B}_{1i}$ in $y_1, y_a, y_b, \cdots, y_f, y_{s+1}, \cdots, y_n$ and their derivatives exceed $p+p_{1a}+p_{1b}+\cdots+p_{1f}$. Since Q_{1i} is of degree p it follows that the degree of B_{1i}' exceeds $p_{1a}+p_{1b}+\cdots+p_{1f}$. Since in addition B_{1i}' is free of y_1 and its derivatives we have our result. To complete the task of showing that G_1' satisfies our hypothesis for the unknowns $y_2, \cdots, y_n$ we must also dispose of the requirement that the degree of each B_{1i}' in $y_2, \cdots, y_n$ exceed $p_{12}+p_{13}+\cdots+p_{1s}$. Of the original $\bar{B}_{1i}$ the only ones which need not meet this requirement are those divisible by $y_2^{p_{12}} \cdots y_s^{p_{1s}}$. These have been removed from consideration by our choice of λ_1. Thus, referring to the argument just given, if $(a, b, \cdots, f)$ coincides with $(2, 3, \cdots, s)$ the alternative (a) is excluded; alternative (b) is what was to have been established.

50. Proceeding in this way we obtain a form G_j' for each F_j whose first term contains y_1 effectively. We have

$$G_j' = \lambda_j y_{i_2}^{p_{j2}} y_{i_3}^{p_{j3}} \cdots y_{i_s}^{p_{js}} + \sum_i u_{ji}' B_{ji}'$$

where $(i_2, i_3, \cdots, i_s)$ is a subset of $s-1$ distinct numbers of the set $(2, 3, \cdots, n)$ and where each B'_{ji} is a power product in $y_2, y_3, \cdots, y_n$ and their derivatives which dominates the first term of G'_j in the required way. Each G'_j goes over into G_j when λ_j is replaced by $\lambda y_1^p + \sum u_{jk}L_{jk}$ and the u'_{ji} are replaced by $u_{ji}Q_{ji}$. Let the number of these forms be q_1.

51. We now introduce new unknowns λ'' and u''_{ji} and new forms

$$G''_j = \lambda'' y_{i_2}^{p_{ji_2}} \cdots y_{i_s}^{p_{ji_s}} + \sum u''_{ji}B_{ji}, \qquad j = 1, \cdots, q_1.$$

When in G''_j the unknown λ'' is replaced by $\lambda_1\lambda_2 \cdots \lambda_{q_1}$ and each u''_{ji} by $\lambda_1 \cdots \lambda_{j-1}\lambda_{j+1} \cdots \lambda_{q_1}u'_{ji}$ then G''_j goes over into $\lambda_1 \cdots \lambda_{j-1}\lambda_{j+1} \cdots \lambda_{q_1}G'_j$.

Let Σ'' be the differential ideal generated by the G''_j. Under the terms of the induction hypothesis we conclude that there is a form $D'' = \lambda''^{w_1} + H''_1$ and an integer t_1, such that for every form $y_{i_2}y_{i_3} \cdots y_{i_s}$ we have

$$(2.8) \qquad (y_{i_1}y_{i_2} \cdots y_{i_s})^{t_1}D''_1 \equiv 0 \ [\Sigma''].$$

H''_1 vanishes for $y_i = 0$, $i = 1, \cdots, n$, which is to say that every term of H''_1 contains some y_{ij} effectively. H''_1 is homogeneous of degree w_1 in λ'' and the u''_{ji}. When λ'' and the u''_{ji} are replaced in D''_1 by the forms indicated above, D''_1 goes over into a form

$$D'_1 = (\lambda_1\lambda_2 \cdots \lambda_{q_1})^{w_1} + H'_1.$$

D'_1 is a form in $y_1, \cdots, y_n, \lambda_1, \cdots, \lambda_{q_1}$ and the u'_{ji}. It is homogeneous of degree q_1w_1 in the λ'_i and the u'_{ji}. Now in D'_1 let λ'_j be replaced by $\lambda y_1^p + \sum u_{jk}L_{jk}$ and u'_{ji} by $u_{ji}Q_{ji}$. Let $p_1 = pw_1q_1$. Then D'_1 goes over into a form

$$D_1 = \lambda^{q_1w_1}y_1^{p_1} + H_1.$$

D_1 is a form in $y_1, \cdots, y_n, \lambda$ and the u_{ji}. It is homogeneous of degree q_1w_1 in λ and the u_{ji}. We are going to show that the degree in the y_{ij} of each term of H_1 exceeds p_1. To this end we consider two types of terms of H_1, those arising from H'_1 and those from $L = (\lambda_1 \cdots \lambda_{q_1})^{w_1}$. Each λ_j was replaced by $\lambda y_1^p + \sum u_{jk}L_{jk}$ where the degree of each L_{jk} in $y_1, \cdots, y_n$ and their derivatives exceeded p. Therefore the degree in the y_{ij} of every term of L except $\lambda^{q_1w_1}y_1^{p_1}$ exceeds p_1. This accounts for terms of H_1 arising from L. As for those arising from H'_1 recall that H'_1 is homogeneous of degree q_1w_1 in the λ_i and the u'_{ji}. Since each λ_i contributes at least p to the degree of H_1 in the y_{ij} and each u'_{ji} contributes exactly p to this degree, the λ_i and u'_{ji} contribute at least $p_1 = pw_1 = q_1$. Because each term of H'_1 was of positive degree in the y_{ij} it follows that the terms of H_1 arising from H'_1 also have a degree in the y_{ij} which exceeds p_1. This verifies our assertion about D_1.

After these replacements are made, equation (2.8) becomes

$$(y_{i_2}y_{i_3} \cdots y_{i_s})^{t_1}D_1 \equiv 0 \ [\Sigma].$$

Consequently, for every $V_j = y_{i_1} y_{i_2} \cdots y_{i_r}$,

$$V_j^{t_1} D_1 \equiv 0 \ [\Sigma].$$

By singling out all of the original F_j whose first terms contain effectively y_k, $k = 2, 3, \cdots, n$, and repeating the above procedure for each k, we obtain finally n forms

$$D_k = \lambda^{q_k w_k} y_k^{p_k} + H_k, \qquad k = 1, \cdots, n,$$

where H_k is homogeneous of degree $q_k w_k$ in λ and the u_{ji} and every term of H_k contains a power product in the y_{ij} whose degree exceeds p_k. There is an integer t such that

$$(2.9) \qquad V_j^t D_k \equiv 0 \ [\Sigma], \qquad k = 1, 2, \cdots, n; j = 1, 2, \cdots, q.$$

Let w be the maximum of the numbers $w_k q_k$ and let each D_k be multiplied by λ^{w-w_k}. Using the same symbols to denote the modified forms, we see that equations (2.9) still hold, and that now the D_k are homogeneous of degree w in λ and the u_{ji}. What we have accomplished by this alteration is to obtain a set of forms $D_k = \lambda^w y_k^{p_k} + H_k$, $k = 1, \cdots, n$, to which we may apply our result for the case $s = 1$. It follows that there is a form $D = \lambda^e + H$ and an integer a such that

$$(2.10) \qquad y_i^a D \equiv 0 \ [D_1, D_2, \cdots, D_n], \qquad i = 1, 2, \cdots, n.$$

Every term of H contains some y_{ij} effectively, and H is homogeneous of degree e in λ and the u_{ji}. (Actually the conclusion that the case $s = 1$ entitles us to draw is that $D = (\lambda^w)^b + H$ where H is homogeneous of degree b in λ^w and the coefficients of the H_k. Since these coefficients are themselves homogeneous of degree w in λ and the u_{ji}, the above conclusion is justified.)

52. We now show that there is an integer t such that

$$V_j^t D \equiv 0 \ [\Sigma], \qquad j = 1, 2, \cdots, q.$$

This will complete the proof, since D meets all our other requirements.

We know that for each D_k there is a power of V_j such that its product with D_k is in Σ. This is likewise true of any derivative of the D_k. We chose the integer h sufficiently large so that the product of V_j with any D_k or with any derivative of a D_k which appears effectively in the right member of some congruence (2.10) shall be in Σ. A single h serves for all j. Clearly

$$V_j^h y_i^a D \equiv 0 \ [\Sigma], \qquad j = 1, \cdots, q; i = 1, \cdots, n.$$

The integer $h + a$ thus has the required property; for every j, $V_j^{h+a} D$ is in Σ.

AN APPROXIMATION THEOREM

53. Our next theorem is an application of Theorem 2.4. Its statement and proof presuppose Raudenbush's theory of perfect differential ideals. The coefficient domain is an arbitrary differential field of characteristic zero.

THEOREM 2.5. *Let Σ be a prime ideal of forms in $y_1, \cdots, y_n$ which does not contain $V = y_1y_2 \cdots y_n$ and which is such that every form of Σ vanishes for $y_i = 0$, $i = 1, \cdots, n$. Let r be a positive integer and let $u_1, \cdots, u_n$ be unknowns. If in each form of Σ, y_{ij} is replaced by $(u_i^r)_j$, $i = 1, \cdots, n$, Σ goes over into a system σ of forms in $u_1, \cdots, u_n$. Let $\{\sigma\}$ be the perfect ideal generated by σ and let it be the intersection of prime ideals $\Omega_1, \cdots, \Omega_e$. Then there is an Ω_i which does not contain $W = u_1u_2 \cdots u_n$ and every form of Ω_i vanishes for $u_i = 0$, $i = 1, \cdots, n$.*

What this amounts to in the theory of differential equations is that the approximation theorem which holds for $r = 1$ holds for any positive integral value of r. In this form our theorem has been established by Ritt([12]). What we shall prove is thus the abstract counterpart of Ritt's approximation theorem relative to the rth roots of the functions constituting a solution of an irreducible system of differential equations. Our proof is indirect. We assume the theorem false and force a contradiction.

54. If the theorem were false, every Ω_i which did not contain W would contain a form $1 + B_i$ where B_i vanishes for $u_i = 0, i = 1, \cdots, n$. Clearly W is not in each Ω_i for then V would be in Σ. Our assumption that the theorem is false implies that there actually are such forms $1 + B_i$. Let their product be $1 + B$. B vanishes for $u_i = 0$, $i = 1, \cdots, n$, and $W(1 + B)$ is in $\{\sigma\}$. Let $F = W(1 + B)$. There is an integer s such that $F^s \equiv 0$ $[\sigma]$. We work back from F^s to a form of Σ.

The forms of σ were obtained from those of Σ by the transformation $y_{ij} = (u_i^r)_j$. Thus while σ is not a different ideal it is closed with respect to differentiation. The inverse of the above transformation may be obtained from the formulas $u_{ij} = ((u_iy_{i1})/(ry_i))_{j-1}$ where the subscript outside the parentheses denotes differentiation. These formulas show that u_{ij}, $j = 1, 2, 3, \cdots$, may be expressed as the product of u_i with a polynomial (rational coefficients) in y_{i1}/y_i and its first $j - 1$ derivatives. Each term of these expressions for the u_{ij} is the quotient of a polynomial in u_i and the y_{ij} by a power of y_i. The total degree of the numerator in u_i and y_{ij} exceeds the degree of the denominator.

55. We examine the effect of making the above replacements for the derivatives of the u_i in a form in the u_i no term of which is free of all the u_{ij}. We obtain a rational function of the u_i and the y_{ij} whose least common denominator is a power product $y_1^{p_1}y_2^{p_2} \cdots y_n^{p_n}$. When the rational function is written in the form $P/(y_1^{p_1}y_2^{p_2} \cdots y_n^{p_n})$ with P a polynomial in the u_i and the

([12]) OCP, pp. 7–14.

y_{ij}, then for each i for which $p_i \neq 0$ each term of P not divisible by $y_i^{p_i}$ is of degree greater than p_i in u_i, y_i and its derivatives. In addition, the total degree of every term of P in the u_i, y_{ij} exceeds $p_1+p_2+\cdots+p_n$. If, for some k, $p_k=0$, then P may be free of the letters u_k, y_{kj}.

Let us suppose these replacements made in F^s. An expression $W^s(1+T)$ is obtained which involves the u_i and the y_{ij}. The expression T is a rational function whose numerator is a polynomial L in the u_i and the y_{ij} and whose denominator is a power product $y_1^{a_1}y_2^{a_2}\cdots y_n^{a_n}$. The remarks made above about the degree of P hold also for L relative to the exponents $a_1, a_2, \cdots, a_n$. F^s belongs to $[\sigma]$ and is a linear combination of forms of σ whose coefficients are forms in $u_1, u_2, \cdots, u_n$. When the above replacements are made in a form of σ, what is obtained is a form in u_i^r and the y_{ij} which, when u_i^r is replaced by y_i, becomes a form of Σ. The fractions with which we deal are produced only by the coefficients which figure in the linear combination. Let $W^s(1+T)$ be multiplied by the least common denominator $y_1^{b_1}y_2^{b_2}\cdots y_n^{b_n}$ of these coefficients. A form

$$G = W^s(y_1^{b_1}y_2^{b_2}\cdots y_n^{b_n} + K)$$

is obtained. G is a polynomial in the u_i and the y_{ij}. It is a linear combination, with polynomials in these letters for coefficients, of polynomials in u_i^r and the y_{ij}, these latter polynomials having the property that when u_i^r is replaced by y_i in them, forms of Σ are obtained. Each term of K dominates $y_1^{b_1}y_2^{b_2}\cdots y_n^{b_n}$ in the same way that each term of L dominates $y_1^{a_1}y_2^{a_2}\cdots y_n^{a_n}$.

56. Let ω be a primitive rth root of unity. Let $G', G'', \cdots, G^{(m)}$ denote the forms obtained by replacing u_i by $\omega^k u_i$ in G; $k=1, \cdots, r-1$; $i=1, \cdots, n$, the replacements being made independently on all the u_i in all possible ways. The product $GG'\cdots G^{(m)}$ is a form in u_i^r and the y_{ij}. Our final substitution is to replace u_i^r by y_i, $i=1, \cdots, n$, in this product. We denote this form in the y_i by P. We have

$$P = V^e(y_1^{q_1}y_2^{q_2}\cdots y_n^{q_n} + H)$$

where H is a form in the y_{ij} such that the total degree of every one of its terms exceeds $q_1+q_2+\cdots+q_n$. In addition, every term of H not divisible by $y_i^{q_i}$ has a total degree in y_i and its derivatives which exceeds q_i provided that $q_i \neq 0$. Clearly P is in Σ. We are going to show that its presence in Σ contradicts our hypothesis.

Since Σ is prime and does not contain V we must have

$$y_1^{q_1}y_2^{q_2}\cdots y_n^{q_n} + H \equiv 0 \; [\Sigma].$$

Let N denote the form constituting the left member of this congruence. We recall that every form of Σ vanishes for $y_i=0$, $i=1, \cdots, n$, and that H also

has this property. Consequently not all the exponents q_i of the first term of N are zero. Renumbering the letters if necessary, suppose that $q_1, q_2, \cdots, q_t$ are not zero, while $q_{t+1}=q_{t+2}= \cdots =q_n=0$. We have $1 \leqq t \leqq n$. We consider N as a form in $y_1, \cdots, y_t$. It follows from Theorem 2.4 that there is a form $1+R$, R vanishing for $y_i=0$, $i=1, \cdots, t$, and an integer a such that

$$(y_1y_2 \cdots y_t)^a(1 + R) \equiv 0 \; [N].$$

Since N and, consequently, $(y_1y_2 \cdots y_t)^a(1+R)$ are in Σ we have the following contradiction. Σ is prime, so it must contain either $y_1y_2 \cdots y_t$ or $1+R$. By hypothesis Σ does not contain $y_1y_2 \cdots y_t$. By hypothesis every form of Σ vanishes for $y_i=0$, $i=1, \cdots, n$, and since $1+R$ does not have this property it cannot be in Σ.

Part III. The differential ideal generated by uv

The form uv and its derivatives

57. Let u and v be unknowns. We investigate the differential ideal Ω generated by the form $X=uv$. For most of the discussion, the underlying ring will be that of forms in the unknowns u and v whose coefficients are rational numbers. Results obtained under these circumstances carry over readily to more general ones.

Our arguments follow the pattern of those used in the discussion of the ideal generated by y^p. We begin with some conventions concerning power products in u and v and their derivatives. We retain for power products in the u_i alone the definitions concerning weight, degree and order already made for the unknown y; and likewise for power products in the v_i alone. A power product P in both the u_i and the v_i may be written in the form UV where U involves only the u_i and V only the v_i. The *signature* of P is defined as an ordered pair of numbers (d_1, d_2) where d_1 and d_2 are the respective degrees of U and V.

We can without fear of confusion describe as *homogeneous* a form all of whose terms have the same signature. The weight of P is defined as the sum of the weights of U and V. A power product P is defined as *higher* than $P'=U'V'$, and P' as *lower* than P, if (a) U is higher than U' or (b) $U=U'$ and V is higher than V'. Evidently if P is different from P' it must be either higher than P or lower than P'. It is evident that our ordering is transitive. Furthermore, if P is higher than P' and G is any power product, then GP is higher than GP'. A β *term* is defined in the following way. Let $P=UV$ be of signature (d_1, d_2). P is a β term if V effectively contains some v_k with $k<d_1$. This implies of course that $d_1>0$. All other power products are called α *terms*. In particular unity, any power product in the u_i alone and any power product in the v_i alone are α terms. It will be noted that these definitions do not respect the symmetry of Ω.

58. We prove the following lemma.

LEMMA 3.1. *Let d_1, d_2, w be any non-negative integers. Every β term $P=UV$ of signature (d_1, d_2) and weight w is congruent* mod Ω *to a linear combination with rational coefficients of α terms of this weight and signature.*

We need only show that P is either in Ω or is a congruent mod Ω to a linear combination with rational coefficients of power products of signature (d_1, d_2) and weight w, all of which are higher than P. Arguments identical with those used in the proof of Lemma 1.1 show how the proof may then be completed.

59. The proof is by induction on d_1, with d_2 and w arbitrary, starting with $d_1=1$. For this case $P=u_kV$ where $k\geqq 0$ and V involves only the v_i. The fact that P is a β term implies that v is effectively present in P and we have $P=u_kvV'$. If $k=0$ then P is in Ω and requires no further discussion. Assume now that $k>0$ and consider X_k, the kth derivative of $X=uv$. It is a form

$$u_kv+\sum_{i=1}^{k}c_iu_{k-i}v_i$$

where the c_i are positive integers. Thus

$$\begin{aligned}P&=(X_k-\sum c_iu_{k-i}v_i)V'\\&\equiv-(\sum c_iu_{k-i}v_i)V'\ [\Omega].\end{aligned}$$

Evidently each term $u_{k-i}v_iV'$ is higher than P and has the same weight and signature as P, so that the statement is verified for $d_1=1$. Observe that in comparing the terms in the right member of this congruence with P, an examination of their factors in the u_i alone reveals that they are higher than P. We attach this fact to the induction hypothesis and assume that the lemma, with this additional restriction, is valid for all power products of signature (d_1, d_2) and weight w, and d_2 and w arbitrary, provided that d_1 is less than some integer $d>0$. It will be shown that it likewise holds for power products of signature (d, d_2) and weight w.

60. Let P be such a β term. Then P effectively contains some v_k with $k<d$. Let u_r be that derivative of u of highest order which is effectively present in P. If $r=0$, then P is in Ω and needs no further discussion. Assuming $r>0$, write

$$P=u_rv_kU'V'.$$

Now the $(r+k)$th derivative of X is a form

$$X_{r+k}=cu_rv_k+\sum_{i=1}^{k}c_iu_{r+i}v_{k-i}+\sum_{i=1}^{r}c_i'u_{r-i}v_{k+i}$$

where c, the c_i and the c_i' are non-negative integers. Certainly c and the c_i' are different from zero and the c_i are zero only if $k=0$. Thus

$$(3.1)\qquad P \equiv \sum_{i=1}^{k} d_i u_{r+i} v_{k-i} U'V' + \sum_{i=1}^{r} d_i' u_{r-i} v_{k+i} U'V' \ [\Omega]$$

where the d_i', d_i are rational numbers. The terms $u_{r-i}v_{k+i}U'V'$ are all of the same weight and signature as P and are all higher than P in the proper way. Consequently only the terms $u_{r+i}v_{k-i}U'V'$ need be considered. Ignoring u_{r+i} for the moment, consider the terms $v_{k-i}U'V'$. Since U' is of degree $d-1$ and $d>k$ these are all β terms. The induction hypothesis applied to them shows that each is congruent to a linear combination of terms $U''V''$ of the same weight and signature, and with U'' higher than U'. Since U'' has the same degree as U', and U' involves only the letters $u, u_1, \cdots, u_r$, it follows that for some $t<r$ (t is a non-negative integer) the exponent of u_t in U'' exceeds that of u_t in U', while for all non-negative integers $s<t$ the exponents of u_s in U'' and in U' are the same. We see then that each $u_{r+i}U''$ is higher than U. Since in the congruence (3.1) each $u_{r+i}v_{k-i}U'V'$ may be replaced by a linear combination of terms $u_{r+i}U''V''$ of the proper weight and signature, the result follows.

Canonical representations

61. A γ *term* is defined to be a form

$$G = EX_{i_1}X_{i_2}\cdots X_{i_s}$$

where E is an α term and $X_{i_1}X_{i_2}\cdots X_{i_s}$ is any power product in the X_i of positive degree. Let E be of signature (d_1, d_2) and weight w. The signature of G is defined to be (d_1+s, d_2+s) and its weight to be $w+i_1+\cdots+i_s$. G is actually a homogeneous isobaric form of this signature and weight. The following lemma shows that the forms defined here as γ terms are entirely analogous to those so defined relative to the ideal $[y^p]$.

Lemma 3.2. *Let H be any homogeneous isobaric element of Ω. Then H may be expressed as a linear combination with rational coefficients of γ terms all of which we have the same weight and signature as H.*

H is a linear combination with rational coefficients of terms $KX_{i_1}\cdots X_{i_r}$, K being some power product in the u_i and v_i. If K is not an α term Lemma 3.1 asserts the existence of a congruence

$$K \equiv \sum c_i K_i \ [\Omega]$$

where the K_i are α terms and the c_i are constants. This congruence may be written as an equality

$$K = \sum_i c_i K_i + \sum_{i,j} c_{ij} K_{ij} X_j$$

where the c_{ij} are constants and the K_{ij} power products. If the right member of this equation is substituted for K in the term $KX_{i_1}\cdots X_{i_r}$ in ques-

tion, a linear combination of γ terms plus a linear combination of terms $K_{ij}X_jX_{i_1}\cdots X_{i_r}$ is obtained. These terms all have a degree in the X_i which exceeds that of the original one. Repetition of this process a finite number of times in the finitely many terms of H must yield an expression for H of the required nature.

The fundamental lemma

62. The following lemma is the counterpart of Lemma 1.3.

LEMMA 3.3. *Let d_1, d_2, w be non-negative integers. The number of γ terms of signature (d_1, d_2) and weight w does not exceed the number of β terms of this weight and signature.*

The plan of the proof is the same as that of Lemma 1.3. A definite β term of this weight and signature will be assigned to each γ term of this weight and signature in such a way that different β terms are assigned to different γ terms.

63. We consider expressions EX_h of the following description. If E is an α term it may be completely arbitrary. If E is a β term it is restricted by h, the exact statement of the restriction requiring that E be written out explicitly. In this case let

$$E = u_{i_1}^{a_1}\cdots u_{i_r}^{a_r}v_{j_1}^{b_1}\cdots v_{j_s}^{b_s} \tag{3.2}$$

where the a_i and b_i are positive integers. We might state explicitly that the subscripts satisfy the relations $i_1<\cdots<i_r$ and $j_1<\cdots<j_s$. Let t be the smallest integer for which $a_1+a_2+\cdots+a_t>j_1$. Our restriction on E is that $i_t+j_1\geqq h$.

A procedure for associating a β term with such an expression will now be described, the β term to have the same weight and signature as EX_h. The β term will have the general appearance u_av_bE where $a+b=h$. Thus requirements of weight and signature will evidently be met. Some preliminary calculations must be made before the β term can actually be produced.

64. Let E be given by equation (3.2). Let $e_i=a_1+a_2+\cdots+a_i$, $i=1,\cdots,r$. Let $f_1=i_1$ and $f_{2n+1}=i_{n+1}+e_n$, $n=1,\cdots,r-1$. Let $f_{2n}=i_n+e_n$, $n=1,\cdots,r$. Finally let f_{2r+1} be an integer which exceeds both h and f_{2r}. Obviously $f_1<f_2<\cdots<f_{2r+1}$. Then either $h\leqq f_1$ or there is a positive integer c such that $f_c<h\leqq f_{c+1}$. Three cases are treated, depending on the relation h bears to the f_i. The cases are

(i) $h\leqq f_1$. E' is defined by

$$E' = u_hvE.$$

(ii) The integer c mentioned above is odd, say $2d-1$, so that $f_{2d-1}<h\leqq f_{2d}$. Then E' is defined by

$$E' = u_{i_d}v_{h-i_d}E.$$

(iii) The integer c is even, say $2d$, so that $f_{2d} < h < f_{2d+1}$. Then E' is defined by

$$E' = u_{h-e_d} v_{e_d} E.$$

NOTE. We admit the possibility that E contains no u_i effectively so that the quantities e_i cannot be computed. For this case the quantity f_{2r+1} is simply f_1, and the case is covered by (i).

The problem now is to show that these assignments always lead to β terms and that distinct expressions are assigned to distinct β terms. It will first be shown that this procedure always leads to a β term, and then the following characterization of the u_a and v_b used as factors with the E will be obtained. The v_b will be shown to be such that no v_k is effectively present in E' with $k < b$. The u_a will be shown to be such that for no $k < a$ does the degree of E' in $u, u_1, \cdots, u_k$ exceed b. The three cases will be treated separately and for each one the validity of these remarks will be shown.

65. For case (i) the fact that E' contains the factor $u_h v$ proves that E' is a β term. It is evident that E' contains no v_k with $k < b$ because b in this case is zero. Finally the fact that $h \leqq i_1$ and $a = h$ shows that E' contains no u_j effectively with $j < a$.

66. For case (ii) it is desirable to write out the inequality $f_{2d-1} < h \leqq f_{2d}$ in full. It states

$$i_d + a_1 + \cdots + a_{d-1} < h \leqq i_d + a_1 + \cdots + a_d. \tag{3.3}$$

(If d is unity this is to mean $i_1 < h \leqq i_1 + a_1$.) In this case E is multiplied by v_b with $b = h - i_d$. The degree of E' in the u_i exceeds that of E in the u_i and the latter degree is certainly not less than $a_1 + \cdots + a_d$. It is a consequence of (3.3) that $a_1 + \cdots + a_d$ is not less than $h - i_d$ so that the degree of E' in the u_i exceeds $h - i_d$. Then E' is a β term since it contains v_b effectively with b less than the degree of E' in the u_i. To show that E' contains no v_k with $k < b$ observe that then E would also contain this v_k. Such an integer would be less than $a_1 + \cdots + a_d$ and this fact with the supposition $k + i_d < b + i_d = h$ would mean that E did not obey the restriction imposed on it. To show that the factor u_a has the property described, note that it is a consequence of (3.3) that $a_1 + \cdots + a_{d-1} < h - i_d = b$. Since E' is identical with E as far as the letters $u, u_1, \cdots, u_{i_d-1}$ are concerned and a is i_d, this inequality shows that for no $k < a$ does the degree of E' in $u, u_1, \cdots, u_k$ exceed b.

67. We now turn to case (iii). The inequality $f_{2d} < h < f_{2d+1}$ written out in full becomes

$$i_d + a_1 + \cdots + a_d < h \leqq i_{d+1} + a_1 + \cdots + a_d \tag{3.4}$$

(for $d = r$ this is to mean $i_r + a_1 + \cdots + a_r < h$).

The factor v_b used with E is in this case defined by $b = a_1 + \cdots + a_d$. Since the degree of E' in the u_i exceeds that of E, and since this latter degree is at

least b, it follows that E' is a β term because it contains v_b effectively. To show that E' contains no v_k effectively with $k<b$ observe that then E would also contain this v_k. It follows from (3.4) that $i_d+k<h$ and since $b=a_1+\cdots+a_d>k$ the restriction imposed on E could not be satisfied. We now show that u_a is such that for all $k<a$ the degree of E' in $u, u_1, \cdots, u_k$ does not exceed b. This degree is the same as that of E in these letters. The integer a is defined as $h-e_d$. It follows from (3.4) that $i_d<h-e_d\leqq i_{d+1}$. The degree of E' in the letters u_k with $k<a$ is thus $a_1+\cdots+a_d$. This number is precisely b, which verifies the statement.

68. It can now be shown that this procedure assigns distinct β terms E' to distinct expressions EX_h. Let E' be any β term. It must contain some of the v_i effectively. Let b be the smallest subscript for which v_b is effectively present in E'. Since E' is a β term its degree in the u_i exceeds b. Let a be the smallest integer such that the degree of E' in $u, u_1, \cdots, u_a$ exceeds b. Then if EX_h led to E' by the method described above, it must have been that $E'=u_av_bE$, $h=a+b$. Thus given E' there is only one possibility for EX_h.

We need the additional fact that if EX_h determines E' as above, and if g is any non-negative integer such that $g\leqq h$, then $E'X_g$ is also an admissible expression. This means that if k is such that the degree of E' in $u, u_1, \cdots, u_k$ exceeds t and v_t actually appears in E', then $k+t\geqq g$. It has already been shown that a and b are the smallest relevant integers and since $a+b=h$ it follows that $a+b\geqq g$.

69. We are now in a position to describe the way in which β terms may be associated with γ terms. Let $G=E_rX_{i_1}X_{i_2}\cdots X_{i_r}$ be a γ term (it is assumed that $i_1\leqq i_2\leqq\cdots\leqq i_r$ and that $r\geqq 1$). E_r is an α term, so that EX_{i_r} is an expression of the type considered and determines a β term E_{r-1} of the same weight and signature. If $r=1$ this β term is associated with G. If $r>1$ it has been shown that $E_{r-1}X_{i_{r-1}}$ is also an admissible expression and determines a β term E_{r-2}. E_{r-2} has the same weight and signature as $E_rX_{i_r}X_{i_{r-1}}$. If $r=2$ then E_{r-2} is associated with G. If $r>2$ we continue. In this way a sequence of β terms $E_{r-1}, E_{r-2}, \cdots, E_0$ is obtained where each E_{j-1} is the β term determined by $E_jX_{i_j}$. E_0 has the same weight and signature as G. We associate it with G.

It will be shown that if G and G' are γ terms associated with E_0 and E_0', and if $E_0=E_0'$, then G and G' are the same. Let $G'=E_s'X_{j_1}\cdots X_{j_s}$. G' and G are to be considered as identical if and only if $E_r=E_s'$, $r=s$ and $(i_1, \cdots, i_r)=(j_1, \cdots, j_s)$. Suppose $E_0=E_0'$. E_0 was determined by $E_1X_{i_1}$ and E_0' by $E_1'X_{j_1}$. Because of the uniqueness property of the procedure used, it follows that $E_1=E_1'$ and $i_1=j_1$. Suppose $r\leqq s$. Then by reconstructing $E_2X_{i_2}, \cdots, E_rX_{i_r}$ it follows that $E_k=E_k'$ and $i_k=j_k$ for $k=1, \cdots, r$. Now E_r is an α term, so that E_r' is also an α term. The only power product in the sequence $E_0', E_1', \cdots, E_s'$ which is an α term is E_s'. It follows that $s=r$ and thus $G=G'$.

It has been shown that every γ term determines a β term having the same weight and signature and that distinct γ terms determine distinct β terms. This completes the proof.

THE STRUCTURE OF THE IDEAL OF uv

70. We now can prove the following lemma.

LEMMA 3.4. *Let d_1, d_2, w be non-negative integers. Let n_α denote the number of α terms A_i of signature (d_1, d_2) and weight w, let n_β denote the number of β terms of this signature and weight, and let n_γ denote the corresponding number of γ terms G_j. Then $n_\beta = n_\gamma$ and a relation*

$$\sum_{i=1}^{n_\alpha} a_i A_i + \sum_{j=1}^{n_\gamma} g_j G_j = 0$$

where the a_i and g_j are rational numbers implies that all the a_i and g_j are zero.

The proof of this lemma is identical with that of Lemma 1.4, so that no further argument will be given. It will be noted that if the a_i and g_j are elements of any differential domain of integrity which contains the rational numbers and over which u and v are unknowns, the same conclusion can be drawn.

71. The above lemmas combine to yield the following theorem.

THEOREM 3.1. *Let $\mathfrak{J}$ be any differential domain of integrity which contains the rational numbers. Let F be any differential polynomial in the unknowns u and v. Then F is expressible in the form*

$$F = \sum a_i A_i + \sum g_j G_j, \qquad a_i, g_i \in \mathfrak{J},$$

where the A_i are α terms and the G are γ terms. For each F there is only one such expression.

For the proof of this theorem the reader is referred to the proof of Theorem 1.1.

COROLLARY. *Let d_1, d_2, w be as in the statement of Lemma 3.4. Let the $\mathfrak{J}$ considered above be a field. Then the number of linearly independent* (mod Ω) *forms with coefficients in $\mathfrak{J}$ which are homogeneous and isobaric of this signature and weight is n_α.*

COROLLARY. *No linear combination of α terms with coefficients in $\mathfrak{J}$ is in Ω.*

INDECOMPOSABILITY OF THE IDEAL OF uv

72. The preceding work enables us to indicate a striking difference between ordinary (algebraic) ideals of polynomials in a finite number of unknowns and differential ideals of such polynomials. This difference is manifested by the differential ideal of uv, as we proceed to show. The mani-

fold of this ideal is reducible into the union of two irreducible manifolds, namely $u=0$ with v arbitrary and $v=0$ with u arbitrary. Nonetheless we are going to show that the differential ideal $[uv]$ has no representation as the intersection or product of two differential ideals whose manifolds are respectively the first and the second just described. Let Σ_1 be any differential ideal of differential polynomials in the unknowns u and v whose manifold is $u=0$ with v arbitrary. Then[13] Σ_1 contains some power of u, say u^r. Again if Σ_2 is a similar differential ideal whose manifold is $v=0$ with u arbitrary, then some power v^t belongs to Σ_2. Suppose that the ideal $[uv]$ had a representation as the intersection or the product of Σ_1 and Σ_2. Since Σ_2 contains v^t, it contains some power v_r^s, and the form $u^r v_r^s$ is in the product and intersection of Σ_1 and Σ_2. This form is an α term and is thus not in $[uv]$. This proves our contention.

The power products in the ideal of uv

73. Theorem 3.2. *Let d_1, d_2, w, n_α be as in Lemma 3.4. A necessary and sufficient condition that $n_\alpha > 0$ is that $w \geqq d_1 d_2$.*

This is equivalent to the assertion that *every power product of signature (d_1, d_2) and weight $w < d_1 d_2$ is in Ω and not every power product of this signature and weight $w \geqq d_1 d_2$ is in Ω.*

We need only investigate the circumstances under which α terms exist. Let $w = d_1 d_2 + h$ with $h \geqq 0$. Then $u^{d_1} v_{d_1}^{d_2 - 1} v_{d_1 + h}$ is an α term of signature (d_1, d_2) and weight w. This disposes of the sufficiency condition. We now show that there are no α terms of signature (d_1, d_2) and weight less than $d_1 d_2$. If $d_1 d_2 = 0$, there are no power products with this property and certainly no α terms. If $d_1 d_2 > 0$, such an α term would have to be such that every v_k effectively present in it would have a subscript not less than d_1. Since it must have d_2 such letters v_k its weight is at least $d_1 d_2$.

74. It might be pointed out that this discussion of Ω applies to more general ideals, in that the X_i need not be the ith derivative of uv. If each X_i is a homogeneous isobaric polynomial in the u_i and v_i of signature (1, 1) and weight i, and is such that every term of this signature and weight is present in X_i with a nonzero coefficient, then the whole discussion applies verbatim. The coefficients of the X_i must be confined to some field and may otherwise be arbitrary.

(13) Raudenbush, loc. cit.

Columbia University,
New York, N. Y.

TOPOLOGY IN LATTICES

BY

ORRIN FRINK, JR.

1. **Introduction.** Many mathematical systems are at the same time lattices and topological spaces. It is natural to inquire whether the topology in such systems is definable in terms of the order relation alone. Sequential topologies of this kind are well known. For example, in the case of the Boolean algebra M/N of measurable sets modulo null sets, and of continuous geometries, the usual topology is that of a metric space, distance being defined in terms of a modular functional [5, pp. 70, 100]. In the theory of partially ordered linear spaces, the notions of sequential order convergence and star convergence, defined by Garrett Birkhoff and Kantorovitch [5, p. 49; 11] in terms of the order relation, are of importance.

However, the topology of many important systems cannot be defined in terms of the convergence of sequences. In this paper, various nonsequential topologies are studied which are definable in terms of the order relation in a lattice. One of these is the topology of Moore-Smith order convergence of directed sets, introduced by Garrett Birkhoff [5, 7]. Another is the *interval* topology, which results on taking the closed intervals of the lattice as a subbasis for the closed sets of the topology. It is shown that in a lattice which is the direct product of chains, these two topologies are equivalent. With respect to its interval topology, any complete lattice is bicompact, but this is not true with the Moore-Smith topology.

Modifications of these topologies are introduced for the complete lattice of all closed sets of a topological space, and the relation of these topologies to those of transformation spaces is considered.

2. **Definitions.** Lattice meet, join, and inclusion will be denoted by $x\cap y$, $x\cup y$, and $x\leqq y$. A system of elements $\{x_a\}$, not necessarily distinct, is called a *directed set* if the subscripts are partially ordered in such a way that given any two subscripts a and b, there exists a third subscript c such that $a\leqq c$ and $b\leqq c$. By a *residual* set of $\{x_a\}$ is meant the set of all x_a such that $a\geqq b$, for some b. A set A of elements of $\{x_a\}$ is said to be *cofinal* if, given any index b, there exists an element x_a of A such that $a\geqq b$. A property is said to hold *residually* or *cofinally* if it holds for a residual or cofinal set of $\{x_a\}$.

By a *closed interval* of a lattice is meant either the entire lattice or a set of elements of one of the three types: (i) all $x\geqq a$, (ii) all $x\leqq b$, (iii) all x such that $a\leqq x\leqq b$. A collection K of closed sets of a space is said to be a *basis* if every closed set of the space is an intersection of sets of K, and K is said to be a *subbasis* if finite unions of sets of K form a basis. The notions of a basis

Presented to the Society September 5, 1941; received by the editors in revised form April 1, 1941.

and a subbasis for the open sets of a space are defined dually. The *interval topology* of a lattice is that which results on taking the closed intervals of the lattice as a subbasis for the closed sets of the space.

THEOREM 1. *Any lattice is a T_1 space with respect to its interval topology.*

Proof. A set made up of a single element a is a closed set, since it is a closed interval consisting of all elements x such that $a \leqq x \leqq a$. The other conditions for a T_1 space hold automatically.

A directed set $\{x_a\}$ of elements of a lattice L is said to *converge* to an element x of L in the *Moore-Smith order topology* if there exist monotonic directed sets $\{u_a\}$ and $\{v_a\}$ such that (1) $u_a \leqq x_a \leqq v_a$, (2) $\sup u_a = x = \inf v_a$, and (3) if $a \leqq b$, then $u_a \leqq u_b$, and $v_a \geqq v_b$. This notion of convergence in a lattice, denoted by $x_a \to x$, was introduced by Garrett Birkhoff [5, 7]. If the suprema and infima involved exist, as they will in a complete lattice, we may also define

$$\limsup x_a = \inf_b \left(\sup_{a \geqq b} x_a\right), \qquad \liminf x_a = \sup_b \left(\inf_{a \geqq b} x_a\right).$$

It is easy to see that $\liminf x_a \leqq \limsup x_a$, and $x_a \to x$ in the Moore-Smith order topology if and only if the two are equal.

The *closure* $\overline{A}$ of a set A of elements of a lattice L in terms of Moore-Smith convergence is defined as follows. An element x of L is in $\overline{A}$ if and only if there exists a directed set of elements of A converging to x. It can be verified that this closure operation has the two properties: (1) $\overline{A \cup B} = \overline{A} \cup \overline{B}$, (2) a set made up of a single element is closed. On the other hand it is not always true in the Moore-Smith topology that closures are closed, that is, that $\overline{\overline{A}} = \overline{A}$. Hence a lattice with this topology is not necessarily a T_1 space.

Two methods of topologizing the same set of elements will be called *equivalent* if they lead to the same definition of closure. It will be shown later by an example that the Moore-Smith order topology of a lattice is not always equivalent to the interval topology.

THEOREM 2. *The lattice operations $x \cap y$ and $x \cup y$ of a distributive lattice are continuous in the Moore-Smith order topology.*

Proof. As Garrett Birkhoff has remarked, it is sufficient to show that the operations are continuous in the variables separately [5, p. 30]. Suppose that $x_a \to x$, where $u_a \leqq x_a \leqq v_a$, u_a and v_a are monotonic, and $\sup u_a = x = \inf v_a$. Then if y is any element of the lattice, $u_a \cap y \leqq x_a \cap y \leqq v_a \cap y$, and it must be shown that $\sup (u_a \cap y) = (\sup u_a) \cap y$, and $\inf (v_a \cap y) = (\inf v_a) \cap y$. The second of these equalities holds in any lattice, and the first is a form of the infinite distributive law. The other half of the theorem is proved dually.

3. **Topology of chains.** By a *chain* is meant a linearly ordered set (sometimes called a simply ordered set). It is natural to define the topology of a

chain by taking the open intervals, consisting of elements x such that $a<x$, or $x<b$, or $a<x<b$, as a basis for the open sets of the space. This gives what Garrett Birkhoff calls the *intrinsic* topology of the chain.

THEOREM 3. *Both the Moore-Smith order topology and the interval topology of a chain are equivalent to the intrinsic topology of the chain.*

Proof. That the interval topology and the intrinsic topology are equivalent follows from the fact that in a chain, the closed and open intervals with one end point are complementary. Suppose now that $x_a \to x$ in the Moore-Smith topology, and u is the lower end point of an open interval containing x. Then $u<x$, and since $\sup u_a = x$, ultimately $u<u_a \leqq x_a$. Likewise, if v is the upper end point of an open interval containing x, ultimately $x_a \leqq v_a < v$. Hence x_a is ultimately in any open interval containing x. Conversely, if every open interval containing x ultimately contains x_a, it follows that $\liminf x_a = x = \limsup x_a$, and consequently $x_a \to x$.

It should be noted that the *sequential* order topology of a chain is not always equivalent to the intrinsic topology. For example, in the chain of all ordinal numbers less than or equal to Ω, the first noncountable ordinal, no sequence of ordinals less than Ω converges to Ω, although some directed sets of such ordinals do.

4. **Cartesian products and direct products.** There are several equivalent methods of defining the cartesian product P of a collection of non-empty topological spaces (X_a). The points x of P consist of selections (x_a) of one point x_a, called the a-coordinate of x, from each space X_a. The topology of P is assigned by taking as a subbasis for the neighborhoods of a point x of P, the collection of all sets of points of P whose a-coordinate is in some neighborhood U_a of the a-coordinate x_a of x. A basis for the neighborhoods of a point x consists of all finite intersections of such subbasic neighborhoods.

If the topology of the spaces (X_a) is given by the convergence of directed sets, an equivalent topology of P results on defining convergence of directed sets in P to mean coordinatewise convergence of the coordinates in the spaces X_a [15, p. 71]. Still a third equivalent topology, if the spaces X_a are T_1 spaces, is obtained by taking as a subbasis for the closed sets of P the collection of all sets of points x of P such that for all a, the a-coordinate of x lies in a subbasic closed set of the space X_a.

By the *direct product* L of a collection of lattices (L_a) is meant the lattice whose elements x are selections (x_a) of one element x_a, called the a-coordinate of x, from each lattice L_a. The elements x and y of L are ordered by the rule that $x \leqq y$ if and only if $x_a \leqq y_a$ for each a, where x_a and y_a are the a-coordinates of x and y. Garrett Birkhoff [5, p. 29] has shown that the *sequential* order topology of the direct product of a *finite* number of lattices is equivalent topologically to the cartesian product of the factors, each with the sequential

order topology. With the interval topology or the Moore-Smith order topology, this result can be extended to the product of an infinite number of factors.

THEOREM 4. *The cartesian product P of any collection of lattices (L_a), each with the interval topology (Moore-Smith order topology), is homeomorphic to the direct product L of these lattices, also topologized by the interval topology (Moore-Smith order topology).*

Proof. Since P and L have the same elements, it remains to show that their topologies are equivalent. For the interval topology this follows from the fact that the same system of closed sets can be taken as a subbasis for both P and L, namely the system of all sets of elements x of P or L whose a-coordinate x_a is in a fixed closed interval of L_a for each a.

In the case of the Moore-Smith order topology, the result is most easily seen by using the definition of cartesian product in terms of coordinatewise convergence of directed sets. The theorem then follows from the fact that betweenness, monotonicity, and suprema and infima are defined coordinatewise in the direct product of lattices.

5. **Some examples.** It follows from Theorems 3 and 4 that the interval topology and Moore-Smith order topology are equivalent, in lattices which are direct products of chains, both to each other and to the cartesian product topology based on the intrinsic topologies of the chains. Many important lattices are of this type.

Let R stand for the lattice of all real numbers ordered by magnitude, and K for the lattice of all complex numbers, ordered as follows: $a+bi \leqq c+di$ if and only if $a \leqq c$ and $b \leqq d$. Then R is a chain and K is lattice-isomorphic with R^2. The usual topology of R and K is clearly equivalent to the interval topology and the Moore-Smith order topology. If a is any cardinal number, then the generalized euclidean or cartesian spaces R^a and K^a are partially ordered linear topological spaces isomorphic with the space of all real- or complex-valued functions defined over a set A of cardinal number a [2, 16]. Considered as lattices, these spaces are isomorphic to the direct product of a lattices, each isomorphic to R or K.

THEOREM 5. *The topology of the generalized euclidean spaces R^a and K^a, considered as cartesian products of a spaces each homeomorphic to R or K, is equivalent to both the interval topology and the Moore-Smith order topology of these spaces, considered as lattices.*

This follows from Theorems 3 and 4. The topology in question is, of course, that of pointwise convergence of directed sets of functions. In many subspaces of R^a and K^a, the two lattice topologies are also of importance. If the elements of R^a are represented as real-valued functions $f(x)$ defined over a set A of cardinal number a, then the subspace consisting of all elements $f(x)$ such that $|f(x)| \leqq m$ is the so-called *Tychonoff cube* J^a, the cartesian product of a intervals J, each of the form $[-m, m]$.

Theorem 6. *The interval topology and the Moore-Smith order topology of the Tychonoff cube J^a, considered as a complete distributive lattice, are equivalent to the ordinary cartesian product topology of J^a.*

This follows from Theorems 3 and 4. It also follows that the space of all elements $f(x)$ of R^a such that $g(x) \leqq f(x) \leqq h(x)$, for $g(x)$ and $h(x)$ fixed, is a bicompact space homeomorphic to J^a with respect to the lattice topologies. There are other subspaces of R^a and K^a in which the lattice topologies are significant. These include the spaces m and c of bounded and convergent sequences, respectively, and M and C of bounded and continuous functions. In certain cases the lattice topologies are equivalent to the weak neighborhood topology of these spaces [1, 10].

6. **Lattices of sets and Boolean algebras.** An important subspace of R^a is the space 2^a of all characteristic functions $f(x)$, that is, of functions taking only the values 0 and 1 over a set A of cardinal number a. The space 2^a is homeomorphic and lattice-isomorphic to the complete Boolean algebra of all subsets of the set A, and also to the direct product of a two-element Boolean algebras. The usual topology of the space 2^a is that defined by Stone [14], which is equivalent to its relative topology as a subspace of the Tychonoff cube J^a. With this topology 2^a is zero-dimensional and bicompact.

Theorem 7. *The interval topology and the Moore-Smith order topology of the complete atomic Boolean algebra 2^a are both equivalent to the Stone topology.*

This follows from Theorem 4. For the case of the Moore-Smith topology it was first proved by Tukey [15, p. 77].

Since 2^a is a complete lattice, its Moore-Smith order topology may be defined in terms of lim sup x_α and lim inf x_α, where $\{x_\alpha\}$ is a directed set of elements of 2^a. Since lim sup x_α consists of all points cofinally in x_α, and lim inf x_α of all points ultimately (residually) in x_α, these notions are a direct generalization of the notions of complete and restricted limit of a sequence of sets, familiar in set theory.

Since any Boolean algebra may be imbedded in an algebra of the form 2^a [13], though only with preservation of finite sums and products, it might be expected that the lattice topologies of a Boolean algebra can be obtained by topological relativization from those of the enveloping algebra 2^a, and are consequently zero-dimensional. Examples show that this is not the case. The fact that the space 2^a is zero-dimensional (totally disconnected) is a consequence of the existence of atomic elements, which are not necessarily present in a sub-algebra. It would be interesting to study the lattice topologies of the complete Boolean algebra M/N of measurable sets modulo null sets, and of the complete Boolean algebra of the regular open sets of a topological space. In its metric topology, and in the equivalent sequential order topology, the algebra M/N is a complete metric space, but it is not bicompact. In its interval topology, however, M/N is bicompact, as will be shown.

7. **Bicompactness.** A space in which closures are defined is said to be *bicompact* if there is a point common to the closures of any collection of sets, if any finite number of the sets have a common point. In a T_1 space, an equivalent condition is that there be a point common to all sets of any collection of basic closed sets, any finite number of which have a common point.

THEOREM 8. *A T_1 space T is bicompact, if there exists a subbasis S for the closed sets of T, such that there is a point common to the members of any collection of sets of S, any finite number of which have a common point.*

This was proved by J. W. Alexander [3]. It can also be proved by means of a complicated argument due to H. Wallman [17, pp. 123–124]. The following proof is simpler.

If K is any collection of basic closed sets of the space T having the finite intersection property, extend K to be maximal by Zorn's lemma [15, p. 7] and call the extended collection M. Since S is a subbasis, any set m of the collection M is the union of a finite number of sets s_r of S. At least one of the sets s_r is also a member of M, since M is maximal. The elements s of S which are also in M have the finite intersection property, hence by hypothesis there is a point p common to them all. The point p is also in every set of K, since each such set is the union of subbasic sets s_r, at least one of which is in M. Hence the space T is bicompact.

THEOREM 9. *Every complete lattice is a bicompact space in its interval topology.*

Proof. A lattice L is said to be *complete* (or continuous) if all subsets of its elements have suprema and infima. Since the maximal and minimal elements I and O exist in a complete lattice, the closed intervals $[x, y]$ with two end points form a subbasis for the closed sets of the interval topology of L. Suppose a collection $\{J_a\}$ of closed intervals $[x_a, y_a]$ is given, every finite number of which have a common element. Then for every pair of indices a and b, we must have $x_a \leqq y_b$, since otherwise the intervals J_a and J_b would have no common element. Since L is complete, sup x_a and inf y_a exist, and sup $x_a \leqq$ inf y_a. Either of these elements is clearly common to all the intervals $\{J_a\}$. Hence, by Theorem 8, L is bicompact. This completes the proof.

8. **Unsolved problem number eleven.** The following is the eleventh of a list of seventeen unsolved problems given in Garrett Birkhoff's *Lattice Theory*, p. 146.

Is every complete lattice topologically bicompact? Is it true that if the intersection of any family of subsets of a complete lattice is void, and if the subsets are closed relative to Moore-Smith convergence, then there exists a finite subfamily having a void intersection?

As we have just seen (Theorem 9), for the interval topology of a complete

lattice, the answer is affirmative. With the Moore-Smith order topology, however, the answer is negative. Consider the complete lattice F of all closed sets of a compact metric space. For simplicity, consider a space S consisting of a point x, and a sequence of points x_n converging to x. The lattice F of closed sets consists of finite sets and of infinite sets containing x. Consider the sets R_m consisting of all x_n such that $n \geqq m$. Each R_m may be thought of as consisting of elements of the lattice F, each made up of a single point.

There is an element common to any finite number of sets R_m, but no element common to them all. It remains to prove that these sets are closed relative to the Moore-Smith order topology of F. Suppose $\{x_\alpha\}$ is any directed set of elements of a particular set R_m, which is not ultimately constant. It is easy to see that lim inf x_α is the empty set. However, lim sup $x_\alpha = \inf v_\alpha$, where v_α is a monotonic directed set of non-empty closed sets of a bicompact space. Hence lim sup x_α is not empty, and the directed set $\{x_\alpha\}$ does not converge. Since the only directed sets of elements of R_m which converge are ultimately constant, R_m is closed in the Moore-Smith topology. Hence F is not bicom pact.

THEOREM 10. *In the complete lattice of all closed sets of a compact metric space, the interval topology is in general distinct from the Moore-Smith order topology.*

THEOREM 11. *There exist complete lattices which are not bicompact with re-respect to the Moore-Smith order topology.*

9. **Comparison of the two lattice topologies.** Of two methods of topologizing the same set of elements, that is, of assigning closures to subsets, one will be called *weaker* than the other if it assigns larger closures to the same sets. In the weaker topology there are fewer closed and open sets. Alexandroff and Hopf [4, p. 62] use the terms *weaker* and *stronger* in exactly the opposite sense. In the case of the lattice F of closed sets of a space, we have seen that the interval topology is sometimes weaker than the Moore-Smith topology. This is true in general.

THEOREM 12. *The interval topology of a lattice is weaker than or equivalent to the Moore-Smith order topology.*

Proof. Suppose the directed set $\{x_\alpha\}$ of elements of a lattice L converges to an element x in the Moore-Smith order topology. In order to prove the theorem, it is sufficient to show that every neighborhood of x in some subbasic system of neighborhoods for the interval topology contains a residual set of $\{x_\alpha\}$. Sets consisting of all elements x not $\leqq b$, or all x not $\geqq c$, being complements of closed intervals, form a subbasis for the neighborhoods of the interval topology. Since $x_\alpha \to x$, there exist monotonic directed sets $\{u_\alpha\}$ and $\{v_\alpha\}$ such that $u_\alpha \leqq x_\alpha \leqq v_\alpha$, and sup $u_\alpha = x = \inf v_\alpha$. Suppose x is not $\leqq b$. Then

it must be shown that ultimately x_a is not $\leqq b$. If this were not true, then $x_a \leqq b$ cofinally. Then since $u_a \leqq x_a$, we would have $u_a \leqq b$ cofinally, and since u_a is monotonic increasing, $u_a \leqq b$ for all a. It would follow that sup $u_a = x \leqq b$, which contradicts the assumption that x is not $\leqq b$. Likewise by duality, if x is not $\geqq c$, then ultimately x_a is not $\geqq c$. This proves the theorem.

10. **Tychonoff's theorem.** As J. W. Alexander has remarked [3], Tychonoff's theorem on the bicompactness of cartesian products is an easy consequence of Theorem 8.

THEOREM 13. *The cartesian product of bicompact T_1 spaces is a bicompact T_1 space.*

Proof. As a subbasis for the closed sets of the product space P we can take the system S of all closed sets of P of the form $F = \Pi_a F_a$, where F_a is an arbitrary closed set of the component space E_a. Since there is clearly a point common to the sets of any subcollection K of S which has the finite intersection property, the space P is bicompact by Theorem 8.

11. **The lattice F of closed sets.** We have seen that in the lattice F of all closed sets of a topological space, the Moore-Smith and interval topologies are not always equivalent. In this case, however, neither of these topologies is satisfactory. In the first place they do not specialize, in the case of metric spaces, to the metric and sequential topologies of Hausdorff [4, p. 111; 9, p. 145]. In the second place they do not specialize by topological relativization, for the subspace consisting of closed sets made up of single points, to the topology of the original space, as would be desirable. However, slight modifications of these two lattice topologies which are also definable in terms of the order relation alone, turn out to be much more useful for the special lattice F, and will now be considered. They will be called the *neighborhood topology* and the *convergence topology*.

12. **The neighborhood topology of F.** An analogue of the interval topology, called the *neighborhood* topology, is defined as follows. Two types of subbasic open sets (neighborhoods) of elements of the lattice F will be considered. Neighborhoods of the *first type* will consist of all elements x of F such that x is not $\leqq a$, where a is any element of F. These are also neighborhoods in the interval topology of F. Neighborhoods of the *second type* will consist of all elements of F not meeting a fixed element b of F, that is, of all x such that $x \cap b = O$, where O is the empty set, that is, the zero element of the lattice. The collection of all finite intersections of neighborhoods of the first or second type is taken as a basis for the open sets of F in the neighborhood topology.

13. **The convergence topology of F.** In the case of the lattice F of closed sets, a more useful limit topology than the Moore-Smith order topology is obtained by retaining the definition of lim sup x_a, while replacing lim inf x_a by a different kind of lower limit. We define LL x_a (which can be read the *lower*

limit of x_a) to be inf (sup A) for all *cofinal* sets A of elements of the directed set $\{x_a\}$. It is clear that LL $x_a \leqq \lim\sup x_a$, since lim sup x_a is by definition inf (sup B) for all *residual* sets B. We write $x_a \rightarrow x$, and say that the directed set $\{x_a\}$ converges to x in the *convergence* topology of F, if and only if LL $x_a = \lim\sup x_a = x$.

Neither the neighborhood topology nor the convergence topology is self-dual. The dual definitions, obtained by reversing the order relation and interchanging sup and inf, can be used to define topologies of the lattice dual to F, consisting of all *open* sets of the space. The definition of the convergence topology was given in the form above in order to show that it depends on the order relation alone. Actually, however, the convergence topology is a simple generalization to directed sets of the notion of topological limit (upper and lower closed limit) of a sequence of sets, due to Hausdorff [9, p. 147].

THEOREM 14. *If $\{x_a\}$ is a directed set of elements of the lattice F of all closed sets of a T_1 space S, then* lim sup x_a *is the set of all points p of S such that every neighborhood of p contains a point of all the elements of some cofinal set of elements of $\{x_a\}$, while* LL x_a *is the set of all points p of S such that every neighborhood of p has a point in common with all the elements of some residual set of elements of $\{x_a\}$.*

Proof. If every neighborhood u of a point p has points in common with every element of a cofinal set of $\{x_a\}$, then every closed set which contains every element of some residual set of $\{x_a\}$, contains p. For if a closed set f contained a residual set of $\{x_a\}$ without containing p, then the complement u of f would be a neighborhood of p not meeting all the elements of any cofinal set. It follows that p is in lim sup x_a.

Conversely, if p is in lim sup x_a, then every closed set which contains a residual set of $\{x_a\}$ contains p. It follows that every neighborhood of p contains a point of every element of some cofinal set of $\{x_a\}$. For if there were a neighborhood u of p which did not, then f, the complement of u, would contain a residual set of $\{x_a\}$ without containing p.

The second part of the theorem states that p is in LL x_a if and only if every neighborhood of p has points in common with all elements of a residual set of $\{x_a\}$. This is proved in the same way as the first part, interchanging the words *cofinal* and *residual*.

THEOREM 15. *In its neighborhood topology, the lattice F of all closed sets* (1) *of a T_1 space is a T_1 space,* (2) *of a regular T_1 space is a Hausdorff space,* (3) *of a bicompact T_1 space is a bicompact T_1 space.*

Proof. If a and b are elements of F and a is not $\leqq b$, then a is a member of the neighborhood of the first type consisting of all elements x of F such that x is not $\leqq b$, but b is not a member of this neighborhood. If p is a point of b but not of a, then b is a member of the neighborhood of the second type

consisting of all elements x of F such that $x \cap p = O$, but a is not a member. This shows that F is a T_1 space in its neighborhood topology.

Now suppose S is a regular T_1 space, and a and b are distinct closed sets of S such that a is not $\leqq b$. If p is a point of a but not of b, there exist disjoint open sets u and v of S, containing p and b, respectively, since S is regular. If u' and v' denote the complements of u and v, respectively, then the neighborhood of the first type, consisting of all elements x of F such that x is not $\leqq u'$, is disjoint from the neighborhood of the second type, consisting of all elements of F such that $x \cap v' = O$. Since these neighborhoods contain a and b, respectively, the lattice F has the Hausdorff separation property.

Finally, suppose S is a bicompact T_1 space. If x and y are elements of the lattice F of closed sets of S, and $x \leqq y$, consider the set $K(x, y)$ consisting of all elements z of F such that $x \cap z \neq O$ and $z \leqq y$. These sets $K(x, y)$, which correspond to the closed intervals of the interval topology, clearly form a subbasis for the closed sets of F in the neighborhood topology. Suppose (K_a) is a collection of such sets $K(x_a, y_a)$, any finite number of which have a common element. It follows that any finite number of the sets (y_a) have a common point. Let y be the intersection of the sets (y_a). Since S is bicompact, y is not empty, and it is clear that we have $x_a \cap y \neq O$ for all a. Hence y is common to all the sets (K_a). It follows from Theorem 8 that F is bicompact.

THEOREM 16. *In the lattice F of all closed sets of a regular T_1 space S, the convergence topology is weaker than or equivalent to the neighborhood topology.*

Proof. Suppose every neighborhood of an element x of F of both the first and second types contains a residual set of the directed set $\{x_a\}$ of elements of F. It must be shown that $x_a \rightarrow x$ in the convergence topology. First it will be shown that $x \leqq \mathrm{LL}\, x_a$. Since every neighborhood of x of the first type contains a residual set of $\{x_a\}$, it follows that if $x_a \leqq u$ cofinally, then $x \leqq u$, where u is an element of F. Hence $x \leqq \sup A$, if A is any cofinal set of elements of $\{x_a\}$. Consequently $x \leqq \mathrm{LL}\, x_a = \inf (\sup A)$ for all A.

Next it will be shown that $x \geqq \lim\sup x_a$. Suppose on the contrary that p is a point of $\lim\sup x_a$, but not of x. Now S is regular; hence there exist disjoint open sets u and v of S containing x and p, respectively. Since every neighborhood of x of the second type contains a residual set of $\{x_a\}$, and $x \cap u' = O$, where u' is the complement of u, then ultimately $x_a \cap u' = O$. Since p is in $\lim\sup x_a$, then cofinally $x_a \cap v \neq O$, by Theorem 14. But this is a contradiction, since $v \leqq u'$.

This proves that $x \geqq \lim\sup x_a$, and it has been shown that $x \leqq \mathrm{LL}\, x_a$. It follows that $x = \lim\sup x_a = \mathrm{LL}\, x_a$, and consequently $x_a \rightarrow x$ in the convergence topology, which was to be proved.

THEOREM 17. *The neighborhood and convergence topologies are equivalent in the lattice F of all closed sets of a bicompact Hausdorff space S.*

Proof. Since a bicompact Hausdorff space is regular, the convergence topology is weaker than or equivalent to the neighborhood topology of F by Theorem 16. It remains to prove the reverse. Suppose then that the directed set $\{x_a\}$ of elements of F converges to x, that is, $x=\text{LL } x_a=\lim\sup x_a$. If $x_a\leqq u$ for a cofinal set A, then $\sup A\leqq u$, hence $\text{LL } x_a=\inf (\sup A)=x\leqq u$. Hence any neighborhood of x of the first type contains a residual set of $\{x_a\}$.

Now suppose some neighborhood of x of the second type does not contain a residual set of $\{x_a\}$, that is, that $x\cap f=O$, but x_b contains a point p_b of f for a cofinal set $\{x_b\}$ of $\{x_a\}$. Since the space S is bicompact, the directed set of points $\{p_b\}$ has a *cluster* point p, such that every neighborhood of p in S contains a cofinal set of $\{p_b\}$ [15, p. 36]. Since f is closed, p is in f. But by Theorem 14, p is in $\lim\sup x_a=x$. This is a contradiction, since $x\cap f=O$. This completes the proof.

THEOREM 18. *If D is any set of points everywhere dense in a T_1 space S, then the collection of all finite subsets of D is everywhere dense in the lattice F of all closed subsets of S, with respect to the neighborhood topology of F.*

Proof. Consider the collection K of all elements of F which are contained in an open set G of S, and have at least one point in common with each of a finite number of open sets G_r of S, each contained in G. It can be seen that the family (K) of all such collections K is a basis (not merely a subbasis) for the open sets of F in the neighborhood topology. But every such collection K contains at least one finite subset of the everywhere dense set D, namely a subset consisting of one point of D selected from each non-empty set G_r. Hence the finite subsets of D are everywhere dense in F, which was to be proved.

The following two theorems, which show the relation of the two topologies of F to other topologies, are easily established and will be stated without proof.

THEOREM 19. *The neighborhood topology and the convergence topology of the lattice F of all closed subsets of a compact metric space S are both equivalent to the Hausdorff metric topology of F.*

THEOREM 20. *The topologies of a Hausdorff space S obtained by topological relativization from the neighborhood topology and the interval topology of the lattice F of all closed sets of S, by considering the points x of S to be elements of F made up of single points, are both equivalent to the original topology of S.*

14. **Transformation spaces.** Since a transformation of a space X into a space Y is determined by its *graph*, which is a set of points in the product space $X\times Y$, any definition of convergence of sets leads to a definition of convergence of transformations.

THEOREM 21. *If $y(x)$ is a continuous transformation of a T_1 space X onto*

a Hausdorff space Y, then the graph G of $y(x)$ is a closed set of the cartesian product space $X\times Y$.

Proof. By the *graph* G of $y(x)$ is meant the set of all points (x, y) of $P=X\times Y$ such that $y=y(x)$. If G is not closed, there is a point (a, b) of P which is in $\overline{G}$, but not in G. Then b is distinct from $y(a)$, and there exist neighborhoods U and V of b and $y(a)$ in Y which are disjoint. Since $y(x)$ is continuous, there is a neighborhood W of a in X such that $y(x)$ is in V if x is in W. The neighborhood $U\times W$ of (a, b) in P contains no points of G, since U and V are disjoint. Hence (a, b) is not in $\overline{G}$, contrary to assumption.

15. **Continuous convergence.** A directed set $\{y_a(x)\}$ of transformations of a space X into a space Y is said to *converge continuously* to the transformation $y(x)$ at a point c of X if for every neighborhood U of $y(c)$ there exists a neighborhood V of c and an index b such that if x is in V and $a\geqq b$, then $y_a(x)$ is in U.

The convergence is said to be *quasi-continuous* at c if for every neighborhood U of $y(c)$ there exists a neighborhood V of c such that if x is any point of V, then there exists an index b such that if $a\geqq b$, then $y_a(x)$ is in U.

THEOREM 22. *If $\{y_a(x)\}$ is a directed set of transformations of a T_1 space X into a regular T_1 space Y which converges pointwise to the transformation $y(x)$, then $y(x)$ is a continuous transformation if and only if the convergence is quasi-continuous at every point of X.*

Proof. If W is a neighborhood of $y(c)$, then there is a neighborhood U of $y(c)$ such that $\overline{U}\leqq W$. If the convergence is quasi-continuous at c, there is a neighborhood V of c such that if x is in V, then there is an index b such that $y_a(x)$ is in U if $a\geqq b$. This is the neighborhood V required by the definition of continuity of the limit transformation $y(x)$. For since the directed set converges pointwise to $y(x)$ at each point x of V, $y_a(x)$ is ultimately in any neighborhood of $y(x)$. But $y_a(x)$ is also ultimately in U, hence $y(x)$ is in $\overline{U}$ and consequently in W. This proves that $y(x)$ is continuous at $x=c$.

Conversely, if $y(x)$ is continuous at $x=c$ and U is any neighborhood of $y(c)$, then there is a neighborhood V of c such that $y(x)$ is in U if x is in V. Then U, being an open set, is also a neighborhood of $y(x)$, and there exists an index b such that $y_a(x)$ is in U if $a\geqq b$, from the convergence. Hence the convergence is quasi-continuous.

COROLLARY. *The limit of a continuously convergent directed set of transformations of a T_1 space into a regular T_1 space, is a continuous transformation.*

Note that it was not assumed that the transformations $y_a(x)$ were themselves continuous. Suppose now that $y_a(x)$ and $y(x)$ are *continuous* transformations of a T_1 space X into a Hausdorff space Y, and that G_a and G are the graphs of these transformations in the product space $P=X\times Y$. Then we have

THEOREM 23. *If the directed set $\{y_a(x)\}$ of transformations converges con-*

tinuously to $y(x)$, *then the directed set* $\{G_a\}$ *of graphs converges to the graph* G *in the convergence topology of the closed sets of the product space* P.

Proof. It is clear that $G \leqq \mathrm{LL}\, G_a$, since every neighborhood of a point of G ultimately contains a point of G_a, from the convergence. Suppose now that the point (c, d) of P is in lim sup G_a, but not in G. Disjoint neighborhoods U and V of the distinct points d and $y(c)$ of Y exist. Since the convergence is continuous, there exists a neighborhood W of c and an index b such that $y_a(x)$ is in V, if x is in W and $a \geqq b$. Since the point (c, d) is in lim sup G_a, there is a point $(x, y_a(x))$ in the neighborhood $U \times W$ of P for some index $a \geqq b$. Since U and V are disjoint, this is a contradiction.

The converse of Theorem 23 is not always true. However, if the spaces X and Y are bicompact Hausdorff spaces, we have

THEOREM 24. *The directed set* $\{y_a(x)\}$ *of transformations from a bicompact Hausdorff space* X *to a bicompact Hausdorff space* Y *converges continuously to the transformation* $y(x)$ *if and only if the directed set of graphs* $\{G_a\}$ *converges to the graph* G *in the convergence topology (or in the equivalent neighborhood topology) of the closed sets of the product space.*

Proof. The necessity of the condition follows from Theorem 23. Suppose $\{y_a(x)\}$ does not converge continuously to $y(x)$ at $x=c$. Then there is a neighborhood U of $y(c)$ such that for every neighborhood V of c, $y_b(x_b)$ is not in U for a cofinal set of indices b, where x_b is in V. The directed set of points $\{x_b, y_b(x_b)\}$ has a cluster point (c, d), since the product space is bicompact. Since U is open, the point d is not in U. But if $G_a \rightarrow G$, the point (c, d), since it is not in G, lies in a neighborhood containing no points of a residual set of $\{G_a\}$, contrary to the assumption that (c, d) is a cluster point. This contradiction proves the sufficiency of the condition.

Theorems 23 and 24 suggest that the topology of continuous convergence might be a suitable one for the semi-group of all continuous transformations of a topological space into itself [8].

REFERENCES

1. L. Alaoglu, *Weak topologies of normed linear spaces*, Annals of Mathematics, (2), vol. 41 (1940), pp. 252–267.

2. J. W. Alexander, *On the connectivity ring of a bicompact space*, Proceedings of the National Academy of Sciences, vol. 22 (1936), pp. 300–303; 381–384.

3. ———, *Ordered sets, complexes, and the problem of bicompactification*, Proceedings of the National Academy of Sciences, vol. 25 (1939), pp. 296–298.

4. P. Alexandroff and H. Hopf, *Topologie* I, Berlin, 1935.

5. Garrett Birkhoff, *Lattice Theory*, American Mathematical Society Colloquium Publications, vol. 25, 1940.

6. ———, *On the structure of abstract algebras*, Proceedings of the Cambridge Philosophical Society, vol. 31 (1935), pp. 433–454.

7. ———, *Moore-Smith convergence in general topology*, Annals of Mathematics, (2), vol. 38 (1937), pp. 39–56.

8. ———, *Hausdorff groupoids*, Annals of Mathematics, (2), vol. 35 (1934), pp. 351–360.

9. F. Hausdorff, *Mengenlehre*, Berlin, 1927.

10. S. Kakutani, *Weak topology, bicompact sets, and the principle of duality*, Proceedings of the Imperial Academy, Tokyo, vol. 16 (1940), pp. 63–67.

11. L. Kantorovitch, *Lineare halbgeordnete Räume*, Recueil Mathématique, (n.s.), vol. 2 (44) (1936), pp. 121–168.

12. R. L. Moore, *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, vol. 13, 1932.

13. M. H. Stone, *The theory of representations for Boolean algebras*, these Transactions, vol. 40 (1936), pp. 37–111.

14. ———, *Applications of the theory of Boolean rings to general topology*, these Transactions, vol. 41 (1937), pp. 375–481.

15. J. W. Tukey, *Convergence and Uniformity in Topology*, Princeton, 1940.

16. A. Tychonoff, *Über einen Funktionenraum*, Mathematische Annalen, vol. 111 (1935), pp. 762–766.

17. H. Wallman, *Lattices and topological spaces*, Annals of Mathematics, (2), vol. 39 (1938), pp. 112–126.

The Pennsylvania State College,
State College, Pa.

OPERATIONS IN BANACH SPACES

BY
MAHLON M. DAY

The starting point for this investigation was an attempt to generalize the well known theorem of Silverman and Toeplitz [25](1) on regularity of a sequence-to-sequence transformation. This theorem may be stated as follows: *If a transformation of sequences* $\{t_n\}$ *of real numbers to sequences* $\{s_m\}$ *is defined from a matrix* $\{a_{mn}\}$, $m, n=1, 2, \cdots$, *by the equations* $s_m=\sum_n a_{mn}t_n$, *the transformation is regular—that is, is defined everywhere and takes every convergent sequence* $\{t_n\}$ *into another convergent sequence with the same limit—if and only if the matrix* $\{a_{mn}\}$ *satisfies the conditions* (a) $\lim_m \sum_n a_{mn}=1$, (b) $\lim_m a_{mn}=0$ *for each n, and* (c) *there is a K such that* $\sum_n |a_{mn}| \leq K$ *for every m.* In the special case under consideration, the fact that regularity implies condition (c) (the non-trivial part of the proof) can be derived from a theorem of Banach [3, p. 80, Theorem 5]: If A and B are Banach spaces, and if U_n, $n=1, 2, \cdots$, are linear operators on A to B, such that $\limsup_n \|U_n(a)\| < \infty$ for each a in A, then $\limsup_n \|U_n\| < \infty$.

If the sequence of integers is replaced by a directed set X, it is known that A, B, X, and U_x can be chosen for which the similar statement relating $\limsup_x \|U_x(a)\|$ and $\limsup_x \|U_x\|$ is false; sections 1–3 of this paper consider these cases in an attempt to solve the problem of boundedness: Characterize those Banach spaces A and B, and directed sets X such that choosing the linear operators U_x on A to B so that $\limsup_x \|U_x(a)\| < \infty$ for each a in A implies that $\limsup_x \|U_x\| < \infty$. Section 1 is a review of pertinent facts about directed sets and convergence (mostly due to Moore and Smith [19], G. Birkhoff [5], and Tukey [27]). Section 2 studies the relations among three topologies in the space of operators on A to B. In §3 the problem of boundedness is studied but not completely solved.

The second part of the paper is concerned with certain special operators on some function spaces. In §4 the space is that of the totally measurable functions on a set Y to a Banach space B; a class of operators on this space is defined in terms of additive, real-valued set-functions and the relations among various topologies in this set of operators is given; this is used in §6 to give a general form to a theorem of Vulich [28]. In §5 the functions studied are the measurable functions on Y to B; the operations on this space are defined in terms of completely additive, limited, set-functions whose values are transfor-

Presented to the Society in three parts, the first under the title *Linear methods of summability* on February 24, 1940; the second under the title *Linear methods of summability.* II on April 27, 1940; the third under the present title on September 5, 1941; received by the editors April 4, 1940, and, in revised form, April 25, 1941.

(1) Numbers in brackets refer to the bibliography at the end of the paper.

mations instead of real numbers. A corollary of the results obtained there is this: *The real-valued, finitely additive set-function* Ψ *has the property that* $\sum_i \Psi(E_i)$ *converges for every sequence* $\{E_i\}$ *of disjoint measurable sets if and only if the total variation of* Ψ *is finite.* This has as a corollary the well known result that a completely additive, real-valued set-function is of bounded variation.

Section 6 contains the applications of these theorems on Banach spaces to the problems which originally started the investigation beginning with two general theorems on regularity conditions. The first includes the Silverman-Toeplitz theorem and many others of the same type; the second generalizes a result of Vulich which says that a transformation defined by a matrix $\{a_{mn}\}$ takes every convergent sequence of points of a Banach space B into another such sequence with the same limit, if and only if the transformation defined by $\{a_{mn}\}$ is regular for real sequences. The section closes with some sample corollaries of these results.

1. **Directed sets and convergence.** This section contains a short discussion of properties of directed sets which will be useful throughout this paper. A non-empty set Y of elements y is called *directed* by a relation $>$ (read "follows" or "succeeds") if the pairs of points y_1, y_2 for which the relation $y_1 > y_2$ holds are subject to the conditions (a) if $y_1 > y_2$ and $y_2 > y_3$ then $y_1 > y_3$ (transitivity) and (b) each pair of points y_1, y_2 in Y has a common successor in Y; that is, there is a y_3 such that $y_3 > y_1$ and $y_3 > y_2$ (composition).

Probably the most used directed set is the set of integers ordered by magnitude. Other examples are (1) the neighborhoods of a point in a topological space ordered by inclusion, and (2) lattices.

A subset Y' is *cofinal* in Y if each y in Y has a successor y' in Y'. It can easily be shown that the cofinal subsets of Y satisfy the following conditions: (3) If y_0 is any element of Y, then the set(2) $\{y \mid y > y_0\}$ is cofinal in Y and its complement $\{y \mid y \not> y_0\}$ is not. (4) If Y_1 is not cofinal in Y and $Y_2 \subset Y_1$, then Y_2 is not cofinal in Y. (5) If Y_1 and Y_2 are not cofinal in Y, then $Y_1 + Y_2$ is not cofinal in Y. (6) If the order relation in a subset of Y is that imposed upon it by the order relation in Y, then every cofinal subset of Y is a directed set.

If f is a real-valued function defined on a directed set Y, let $\limsup_y f(y)$ be the least upper bound of those numbers K for which $\{y \mid f(y) > K\}$ is cofinal in Y; let $\liminf_y f(y) = -\limsup_y (-f(y))$; if $\limsup_y f(y) = \liminf_y f(y)$, then call this common value $\lim_y f(y)$. The Cauchy criterion is a necessary and sufficient condition for existence of $\lim_y f(y)$ and the limit defines an additive, homogeneous functional on those f's for which it is defined.

G. Birkhoff [5] first extended this definition of convergence to a topological space. Let S be a neighborhood space, satisfying, say, the axioms defining a Hausdorff space [2]; if f is a function defined on the directed set Y with

(2) $\in$, $\subset$, and so on will have the usual set-theoretical meanings; $\{y \mid \cdots\}$ will mean the class of those y satisfying the condition following the vertical bar.

values in the space S, $s=\lim_y f(y)$ if and only if for each neighborhood N of s there is a y_N in Y such that $f(y)\in N$ if $y>y_N$. If S is a complete metric space and if the neighborhoods of a point are the spheres about that point, then the Cauchy criterion is again necessary and sufficient for the existence of a limit; if S is a linear space with a uniform topology—that is, a topology in which addition of elements and multiplication of elements by real numbers are continuous operations—then this limit defines an additive, homogeneous function over the linear set of f's on Y to S for which it exists. In particular if S is a Banach space in any of the usual topologies, this is the case.

Tukey [27] has shown the importance of certain directed sets (first used by Moore) and has defined an order relation among directed sets which will be useful in later sections. For any ordinal number $\nu>0$, let D^ν be a set of power $\aleph_\nu$; the *stack* Δ^ν is the directed set whose elements are the finite subsets δ of D^ν, where $\delta>\delta'$ means $\delta\supset\delta'$; D^ν is the *base* of the stack Δ^ν. It is clear that if two stacks have bases of the same power, then there is an isomorphism—that is, a 1-1 order-preserving correspondence—between the two stacks.

A directed set X is a *cofinal part* of a directed set Y if there is an isomorphism between X and a cofinal subset Y' of Y. X and Y are *cofinally similar* (symbol: $X\sim Y$) if there is a directed set Z of which X and Y are both cofinal parts. X follows Y (symbol: $X>Y$) if and only if there exist two functions, h on X to Y and g on Y to X, such that if y is any point of Y and $x>g(y)$, then $h(x)>y$. Tukey showed that $X>Y$ and $Y>X$ if and only if $X\sim Y$, that $>$ is a reflexive and transitive ordering among directed sets, and that cofinal similarity is a reflexive, symmetrical and transitive relation, the equivalence relation associated with $>$.

The reader can easily see that ω, the class of integers ordered by magnitude, is a cofinal part of Δ^0, the stack on a countable base, and that the stack Δ^ν follows every directed set of power less than or equal to $\aleph_\nu$; also $\Delta^\nu>\Delta^\mu$ if and only if $\nu\geqq\mu$, so $\Delta^\nu\sim\Delta^\mu$ if and only if $\nu=\mu$. From this it follows that $\Delta^0>X$ if X is any countable directed set. If $\Delta^0>X$ either X has a last element—that is, an x_0 such that $x_0>x$ for each x in X—or $\Delta^0\sim X$. In all of what follows the trivial case will be explicitly rejected; that is, no directed sets mentioned hereafter will have a last element.

If X is a directed set, let $\lambda(X)$ be the smallest ordinal number μ such that a subset of X of power $\aleph_\mu$ has no upper bound in X; that is, $\lambda(X)$ is the smallest ordinal μ not satisfying the following condition: If $X'\subset X$ and the power of $X'\leqq\aleph_\mu$, there is an x_0 in X such that $x_0>x'$ for every x' in X'. For example, $\lambda(\Delta^\nu)=0$ for every ν; $\lambda(\omega)=0$; $\lambda(\Omega_n)=n$ for any integer n if Ω_n is the set of all ordinals of power less than $\aleph_n$ ordered by magnitude. From the definition it can readily be seen that if $X>Y$, then $\lambda(X)\leqq\lambda(Y)$, so $\lambda(X)$ is invariant under cofinal similarity. The next lemma is useful in §3.

LEMMA 1.1. *If X is a directed set, then $\lambda(X)=0$ if and only if $X>\Delta^0$.*

If $X > \Delta^0$, then $0 \leqq \lambda(X) \leqq \lambda(\Delta^0) = 0$. If $\lambda(X) = 0$, a countable set $\{x_n'\}$ exists with no upper bound; by induction and the composition property a sequence $\{x_n\}$ can be defined so that $x_{n+1} > x_n$ and x_n'; then $\{x_n\}$ has no upper bound and is monotone. Define h on X to ω and g on ω to X by letting $g(n) = x_n$ for each n in ω; $h(x) = n+1$ if $x_n < x \not< x_{n+1}$.

The interested reader can also prove that no X can be chosen for which $\lambda(X) = \omega$; this fact clarifies some steps of the proof of Theorem 3.6.

2. **Neighborhoods and convergence in operator spaces.** This section considers relations among two Banach spaces[3] A and B and the space $\mathfrak{U} = A:B$ of all linear[4] operators defined over all of A with values in B; $\mathfrak{U}$ is also a Banach space if $\|U\| = \sup_{\|a\| \leq 1} \|U(a)\|$ for each U in $\mathfrak{U}$. In the special case in which B is B_0, the set of real numbers, $A:B_0$ is the space A^* of all linear functionals on A. There are three natural ways in which a topology can be imposed on $\mathfrak{U}$; by analogy with the case in which $A = B =$ Hilbert space[5] these will be called norm, s^*, and w^* topologies in $\mathfrak{U}$. It is sufficient (see Wehausen, [29]) to define the neighborhoods of θ[6]; the neighborhoods of the other points of $\mathfrak{U}$ are defined by translating the neighborhoods of θ.

NORM: *For any* $\epsilon > 0$ *let* $N = N(\epsilon) = \{U \mid \|U\| < \epsilon\}$.

S^*: *For any integer* k, *any* $a_1, \cdots, a_k$ *in* A, *and* $\epsilon > 0$ *let* $S = S(a_1, \cdots, a_k; \epsilon) = \{U \mid \|U(a_i)\| < \epsilon$ *for* $i = 1, \cdots, k\}$.

W^*: *For any integer* k, *any* $a_1, \cdots, a_k$ *in* A, *and* $\beta_1, \cdots, \beta_k$ *in* B^* *and any* $\epsilon > 0$ *let* $W = W(a_1, \cdots, a_k; 1, \cdots, k; \epsilon) = \{U \mid |\beta_i(U(a_i))| < \epsilon$ *for* $i = 1, \cdots, k\}$.

The families $\mathfrak{N}$, $\mathfrak{S}$, and $\mathfrak{W}$ of these sets N, S, and W are, respectively, the norm, s^*, and w^* neighborhoods of θ in $\mathfrak{U}$; in the special case $B = B_0$ both s^* and w^* topologies reduce to the ordinary weak* topology in A^*[7]. If X is any directed set and if $U_x \in \mathfrak{U}$, the notations $U_0 = n - \lim_x U_x$, $U_0 = s^* - \lim_x U_x$ and $U_0 = w^* - \lim_x U_x$ mean that U_x converges to U_0 in the corresponding topology.

The first half of the next theorem is used in §3. Two neighborhood systems $\mathfrak{N}'$ and $\mathfrak{N}''$ of θ in $A:B$ will be called equivalent (symbol: $\mathfrak{N}' \simeq \mathfrak{N}''$) if each N' contains an N'' and each N'' an N'. Clearly each W contains an S and each S an N. A Banach space B is called finite-dimensional (symbol: fd) if there exist a finite subset $b_1, \cdots, b_k$ in B such that every b in B is a linear combination of these b_i.

---

(3) A Banach space [see 3] is a complete normed vector space. In all that follows the trivial space consisting of just one point will be ruled out and all spaces considered will be at least one-dimensional.

(4) Linear is used in Banach's sense, to mean additive and continuous.

(5) See J. von Neumann [20], for this case; others who have considered topologies in a Banach space are A. E. Taylor [25], and Alaoglu [1].

(6) θ will be used for the zero element of any linear space under discussion.

(7) See Taylor [25].

THEOREM 2.1. *$\mathfrak{N} \simeq \mathfrak{S}$ if and only if A is* fd; *$\mathfrak{S} \simeq \mathfrak{W}$ if and only if B is* fd; *hence $\mathfrak{N} \simeq \mathfrak{W}$ if and only if both A and B are* fd([8]).

If A is fd, there is a basis $a_1, \cdots, a_k$ of linearly independent points of A with $a = \sum_{i \leq k} t_{ai} a_i$, t_{ai} real, for each a in A. Since every two k-dimensional Banach spaces are isomorphic there is a $K > 0$ such that $\sum_{i \leq k} |t_{ai}| \leq K$ if $\|a\| \leq 1$. Recall that $\|U\| = \sup_{\|a\| \leq 1} \|U(a)\|$; if $U \in S(a_1, \cdots, a_k; \epsilon/K)$, then

$$\|U(a)\| \leq \sum_{i \leq k} |t_{ai}| \, \|U(a_i)\| < \frac{K\epsilon}{K} = \epsilon$$

if $\|a\| \leq 1$. Therefore $N(\epsilon) \supset S(a_1, \cdots, a_k; \epsilon/K)$ and $\mathfrak{N} \simeq \mathfrak{S}$ when A is fd. If A is not fd and $S = S(a_1, \cdots, a_k; \epsilon)$ is any s^* neighborhood of θ, there is an α in A^* such that $\alpha(a_i) = 0$ for each $i \leq k$, while $\|\alpha\| > 0$. If $\|b\| \neq 0$, the element U_n of U defined by $U_n(a) = n\alpha(a)b$ is in S for every n while $\|U_n\| = n\|\alpha\| \, \|b\| \to \infty$ as n increases, so S is not contained in any sphere.

If B is fd, let $\beta_1, \cdots, \beta_q$ and $b_1, \cdots, b_q$ be conjugate bases in B^* and B, respectively; that is, choose them so that $\beta_i(b_i) = 1$, $\beta_i(b_j) = 0$ if $i \neq j$, and $\beta_1, \cdots, \beta_q$ and $b_1, \cdots, b_q$ are linearly independent and are bases in their respective spaces; then any b in B is of the form $\sum_{j \leq q} \beta_j(b) b_j$. If $S = S(a_1, \cdots, a_k; \epsilon)$ is given, the w^* neighborhood for which $|\beta_j(U(a_i))| < \epsilon/(q \sup_j \|b_j\|)$ for every $i \leq k, j \leq q$ lies in S since

$$\|U(a_i)\| = \left\| \sum_{j \leq q} \beta_j(U(a_i)) b_j \right\| < \sum_{j \leq q} |\beta_j(U(a_i))| \, \|b_j\| < \epsilon$$

if $|\beta_j(U(a_i))| < \epsilon/(q \sup_j \|b_j\|)$ for all i, j; hence $\mathfrak{S} \simeq \mathfrak{W}$. If B is not fd and if $W = W(a_1, \cdots, a_k; \beta_1, \cdots, \beta_k; \epsilon)$ is given, there is a point b in B with $\beta_i(b) = 0$ for all $i \leq k$ while $\|b\| > 0$. If $a \in A$, if $\|a\| \neq 0$ and if α is any element of A^* for which $\alpha(a) \neq 0$, each U_n defined by letting $U_n(a') = n\alpha(a')b$ for each a' in A is in W since $\beta_i(U_n(a)) = n\alpha(a)\beta_i(b) = 0$; $\|U_n(a)\| = n|\alpha(a)| \, \|b\| \to \infty$ as n increases so W cannot lie in any $S(a, \epsilon)$ for which $\|a\| \neq 0$.

3. **The boundedness problem.** The theorem of Banach [3, p. 80, Theorem 5] already mentioned asserts that if $\{U_n\}$ is any sequence of elements of $A:B$ such that $\limsup_n \|U_n(a)\| < \infty$ for every a in A, then $\limsup_n \|U_n\| < \infty$. The boundedness problem is to characterize those triples A, B, X such that $\limsup_x \|U_x\| < \infty$ if $\limsup_x \|U_x(a)\| < \infty$ for each a. Some unsettled questions connected with this problem are collected at the end of this section.

Consider the following conditions:

(a) $\limsup_x \|U_x(a)\| < \infty$ for each a in A.

(b) $\lim_x \|U_x(a)\| = 0$ for each a in A.

---

(8) Even in the unit sphere $\mathfrak{U}_1$ in $\mathfrak{U}$, the s^* and w^* topologies are generally different; this can be seen from the result, more general than one of Alaoglu [1], that $\mathfrak{U}_1$ is bicompact in the s^* topology if and only if B is fd, while $\mathfrak{U}_1$ is bicompact in the w^* topology if and only if B is reflexive.

(c) $\limsup_x \|U_x\| = \infty$.
(d) $\lim_x \|U_x\| = \infty$.

The first step in the solution of the boundedness problem is to show that the nature of B is unimportant.

THEOREM 3.1. *If A and X are given and if a B exists such that linear operators U_x can be defined on A to B so as to satisfy any combination of the conditions* (a)–(d), *then for any B', U'_x can be chosen in $A:B'$ to satisfy the same conditions.*

If the linear operators U_x on A to B are given, for each x let β_x be an element of B^* such that $\|\beta_x\| = 1$ and

$$\|U_x\| = \sup_{\|a\| \leq 1} \|U_x(a)\| \leq 2 \sup_{\|a\| \leq 1} |\beta_x(U_x(a))|.$$

Let α_x in A^* be defined by $\alpha_x(a) = \beta_x(U_x(a))$; then $|\alpha_x(a)| \leq \|U_x(a)\|$ while $\|U_x\| < 2\|\alpha_x\|$ so the α_x satisfy those conditions which are satisfied by the U_x. If B' is any other space, let b' be any point in B' for which $\|b'\| = 1$ and define U'_x by the equation $U'_x(a) = \alpha_x(a)b'$; then $\|U'_x\| = \|\alpha_x\|$ and $\|U'_x(a)\| = |\alpha_x(a)|$ so the U'_x have the same properties.

For any combination of the conditions (a)–(d) let $\mathfrak{P}$ with those subscripts be the class of all pairs $[A, X]$, where A is a Banach space and X a directed set, such that α_x in A^* exist satisfying that set of conditions; for example, $\mathfrak{P}_{ac}$ is the set of all $[A, X]$ such that α_x in A^* exist with $\limsup_x |\alpha_x(a)| < \infty$ for each a in A and $\limsup_x \|\alpha_x\| = \infty$. The problem of boundedness is to characterize $\mathfrak{P}_{ac}$; related to this are the problems of characterizing the sets $\mathfrak{P}_{ad}$, $\mathfrak{P}_{bc}$, and $\mathfrak{P}_{bd}$. There are several obvious relations among these classes; $\mathfrak{P}_{bd} \subset \mathfrak{P}_{ad} \subset \mathfrak{P}_{ac}$, $\mathfrak{P}_{bd} \subset \mathfrak{P}_{bc} \subset \mathfrak{P}_{ac}$, and $\mathfrak{P}_d \supset \mathfrak{P}_{ad}$; to be proved later (Theorem 3.7) is the fact that $\mathfrak{P}_{ad} = \mathfrak{P}_{bd}$.

Consider first some "monotony" properties of these sets. A Banach space A' will be called a *linear image* of a Banach space A if there is a linear operator U on A to B whose values fill up B. If U is such an operator and if $A_0 = \{a \mid U(a) = \theta\}$, then A' is isomorphic to the Banach space A/A_0(9).

THEOREM 3.2. *If $[A, Y] \in \mathfrak{P}_{bd}$, $\mathfrak{P}_{bc}$, or $\mathfrak{P}_{ac}$ and $X > Y$, then $[A, X]$ is in the same class. If A' is a linear image of A and if $[A', X] \in \mathfrak{P}_{bd}$, $\mathfrak{P}_{bc}$, or $\mathfrak{P}_{ac}$, then $[A, X]$ is in the same class.*

If $X > Y$, there are functions g on Y to X and h on X to Y such that $h(x) > y$ if $x > g(y)$. If $[A, Y] \in \mathfrak{P}_{bd}$, there exist α_y in A^* such that $\lim_y |\alpha_y(a)| = 0$ for each a in A and $\lim_y \|\alpha_y\| = \infty$. Let $\alpha_x = \alpha_{h(x)}$; then for each a in A and

(9) If A_0 is a closed linear subset of A, the elements of A/A_0 are the cosets $E_a = \{a_1 \mid a - a_1 \in A\}$ where $\|E_a\| = \inf_{a_1 \in E_a} \|a_1\|$; with the usual definitions of the vector operations A/A_0 is a Banach space. If $A_0 = U^{-1}(\theta)$, the elements of A/A_0 are the sets $E = U^{-1}(a')$, a' in A'; the transformation T on A/A_0 to A' defined by $T(U^{-1}(a')) = a'$ is linear and 1-1 so [3, p. 41, Theorem 5] it is an isomorphism.

$\epsilon>0$ there is a y_ϵ such that $|\alpha_y(a)|<\epsilon$ if $y>y_\epsilon$. Let $x_\epsilon=g(y_\epsilon)$; if $x>x_\epsilon$, then $|\alpha_x(a)|=|\alpha_{h(x)}(a)|<\epsilon$ since $h(x)>y_\epsilon$, so $\lim_x|\alpha_x(a)|=0$; similarly $\lim_x\|\alpha_x\|=\infty$.

If $[A, Y]\in\mathfrak{P}_{ac}$ or $\mathfrak{P}_{bc}$, let α_y in A^* have the corresponding properties. Define h_1 on X to Y as follows: Suppose that h_1 is already defined on a subset X' of X so that (1) X' contains every predecessor of each of its elements, (2) for each x' in X' there is a sequence $\{x_n'\}\subset X'$ such that $x_n'>x'$ and $\lim_n\|\alpha_{h_1(x_n')}\|=\infty$, and (3) $h_1(x')>h(x')$ for every x' in X'. If $X'\neq X$, let x be any element of $X-X'$ and let $\{x_n\}$ be any sequence of points of $X-X'$ such that $x_{n+1}>x_n$ for each n while $x_1>x$. Since $\lim\sup_y\|\alpha_y\|=\infty$, for each n there exists a point $h_1(x_n)$ in Y such that $h_1(x_n)>h(x_n)$ and $\|\alpha_{h_1(x_n)}\|>n$. Let $X''=X'+\{x''|$ an n exists for which $x_n>x''\}$; for each x'' in X'' for which $h_1(x'')$ is not already defined let $h_1(x'')=h(x'')$; then h_1 is defined over X'' with the properties (1)–(3). Starting with X' equal to the empty set and applying transfinite induction defines h_1 over all X with the properties (2) and (3). From (3) and a repetition of the argument in the preceding paragraph it follows that the α_x defined by $\alpha_x=\alpha_{h_1(x)}$ satisfy (a) or (b) if the α_y do; (2) implies that α_x satisfy (c).

Suppose that the linear operator U maps A onto all of B, let $A_0=U^{-1}(\theta)$, and construct A/A_0. If $\alpha_x'\in A'^*$, define α_x in A^* by $\alpha_x(a)=\alpha_x'(U(a))$; clearly the α_x satisfy (a) or (b) if the α_x' do. For each x

$$\|\alpha_x\| = \sup_a(|\alpha_x(a)|/\|a\|) = \sup_a(|\alpha_x'(U(a))|/\|a\|).$$

For each $\epsilon>0$ there is an a' of norm one such that $\alpha_x'(a')>\|\alpha_x'\|-\epsilon$; if(9) $E=T^{-1}(a')$, $\|E\|\leqq\|T^{-1}\|$ so there is an a in E of norm less than $\|T^{-1}\|+\epsilon$; hence $\|\alpha_x\|>(\|\alpha_x'\|-\epsilon)/(\|T^{-1}\|+\epsilon)$ for every $\epsilon>0$ or $\|\alpha_x\|\geqq\|\alpha_x'\|/\|T^{-1}\|$. Therefore the α_x satisfy (c) or (d) if the α_x' do.

We now consider a case in which $[A, X]$ can be shown to be in the smallest of these classes.

THEOREM 3.3. *If A is not* fd, *if B is any Banach space, and if $\mathfrak{S}$, the s^* neighborhood system of θ in $A:B$, is directed by the relation $S>S'$ if $S\subset S'$, then $[A, \mathfrak{S}]\in\mathfrak{P}_{bd}$.*

If $S\in\mathfrak{S}$, there is a least integer k such that $S=S(a_1,\cdots,a_k;\epsilon)$. If A is not fd, by Theorem 2.1, S contains a point U_S for which $\|U_S\|>k$; then U_S defined in this way have the properties (b) and (d). If $a\in A$ and $\epsilon>0$ is given, $U_S\in S\subset S(a;\epsilon)$ if $S>S(a;\epsilon)$, so $\|U_S(a)\|<\epsilon$ if $S>S(a;\epsilon)$ or $\lim_S\|U_S(a)\|=0$. If $S=S(a_1,\cdots,a_k;\epsilon)\subset S'=S(a_1',\cdots,a_q';\epsilon')$, then each a_j' is linearly dependent on the a_i. For suppose that some a_j' does not depend on the a_i; then [3, p. 57, lemma] there is an α in A^* such that $\alpha(a_i)=0$ for all i while $\alpha(a_j')=1$. Take $b\neq\theta$ in B and let $U(a)=\alpha(a)b$; then $kU\in S$ for every k, but if $k\|b\|>\epsilon'$, $kU\notin S'$; this contradicts the assumption that $S\subset S'$. If $a_1',\cdots,a_q'$ are

chosen linearly independent, then $S(a_1, \cdots, a_k; \epsilon) \subset S(a_1', \cdots, a_q'; 1/q)$ implies that $k \geqq q$, so $\| U_S \| > q$ if $S > S(a_1', \cdots, a_q'; 1/q)$, and $\lim_S \| U_S \| = \infty$.

Banach's theorem asserts that $[A, \omega] \notin \mathfrak{P}_{ac}$ for any A; a converse of this is contained in the first half of the next theorem.

THEOREM 3.4. *$\Delta^0 > X$ if and only if there is no A for which $[A, X]$ is in $\mathfrak{P}_{ac}$ (or $\mathfrak{P}_{bc}$). A is* fd *if and only if there is no X for which $[A, X]$ is in $\mathfrak{P}_{ac}$ (or $\mathfrak{P}_{bd}$).*

If $[A, X] \in \mathfrak{P}_{ac}$ and $\Delta^0 > X$, then $[A, \omega] \in \mathfrak{P}_{ac}$, by Theorem 3.2; this contradicts Banach's theorem. If $\Delta^0 \not> X$, no countable subset of X is cofinal in X. Let $A = c_0(X, B_0)$(10), where B_0 is the space of real numbers, and define α_x in A^* in a way similar to that used in the proof of Theorem 3.3. Suppose that the α_x have been defined on a subset X' of X to satisfy the conditions (1) X' contains every predecessor of each element of X', (2) for each x' in X', there is a monotone sequence $\{x_n'\} \subset X'$ such that $\lim_n \|\alpha_{x'_n}\| = \infty$, and (3) for each x' in X', $\alpha_{x'}$ is defined by $\alpha_{x'}(a) = k_{x'}a(x')$ for each a in A, and some constant $k_{x'}$. Then take any x not in X' and any monotone sequence $\{x_n\} \subset X - X'$ such that $x_1 > x$ and for each n define α_{x_n} in A^* by $\alpha_{x_n}(a) = na(x_n)$ for each a in A; define $\alpha_{x'}$ in A^* for those predecessors x' of any x_n for which it is not already defined by setting $\alpha_{x'} = \theta$. As before, this and transfinite induction serve to define α_x for every x in X so that $\limsup_x \|\alpha_x\| = \infty$. For each a in A the set $E_a = \{x \mid |a(x)| > 0\}$ is countable, hence not cofinal in X; therefore there is an x_a in X such that no x in E_a follows x_a, so $\lim_x \alpha_x(a) = 0$ for each a since $\alpha_x(a) = 0$ if $x > x_a$. This shows that $[c_0(X, B_0), X] \in \mathfrak{P}_{bc}$ unless $\Delta^0 > X$.

If A is fd, by Theorem 2.1, norm and strong neighborhoods systems are equivalent; using this fact it is clear that no $[A, X] \in \mathfrak{P}_{bc}$. To show that no $[A, X]$ can be in $\mathfrak{P}_{ac}$ requires only an application of the method of proof used in that theorem. If A is not fd, Theorem 3.3 asserts that an X exists with $[A, X]$ in $\mathfrak{P}_{bd}$.

We turn now to a characterization of the nature of $\mathfrak{S}$ considered as a directed set rather than as a neighborhood system.

THEOREM 3.5. *A is not* fd *if and only if for every B there is an ordinal number $\nu > 0$ such that Δ^ν is a cofinal part of the directed set $\mathfrak{S}$ of strong neighborhoods of θ in $A:B$; ν is unique and depends only on A.*

Proof. If A is not fd, let A' be a vector basis in A; that is, a set of points a' of norm one, such that no a' in A' is linearly dependent on any of the others, while every element of A is a linear combination of elements of A'. Let ν be the ordinal for which the power of A' is $\aleph_\nu$. If B is any Banach space, we shall show that Δ^ν with this choice of ν is a cofinal part of $\mathfrak{S}$, the strong neighborhood system in $A:B$; clearly ν does not depend on B but only on A.

Let f be any 1-1 correspondence between the class of neighborhoods $S(a; 1)$, a in A', and D^ν, the base of the stack Δ^ν; extend f to all of Δ^ν by letting

(10) This is defined before Corollary 3.3.

$f(\delta)=S(f(d_1), \cdots, f(d_k); 1/k)$ if $d_1, \cdots, d_k$ are the elements of δ. Then f defines a 1-1 correspondence between Δ^ν and a certain subset $\mathfrak{S}'$ of $\mathfrak{S}$.

$\mathfrak{S}'$ is cofinal in $\mathfrak{S}$, for if $S(a_1, \cdots, a_k; \epsilon)\in\mathfrak{S}$, there exist a_{ij} in A' and real numbers t_{ij} such that $a_i=\sum_{j\leq k_i} t_{ij}a_{ij}$. Then S_1, the neighborhood such that $|U(a_{ij})| < \epsilon/(\sum_{ij}|t_{ij}|)$, is contained in S. If enough additional elements of A' are used to make $1/k$ smaller than $\epsilon/(\sum_{ij}|t_{ij}|)$, S_1 contains some $S'\in\mathfrak{S}'$, so $\mathfrak{S}'$ is cofinal in $\mathfrak{S}$.

f preserves order between $\mathfrak{S}'$ and Δ^ν. Obviously $f(\delta)>f(\delta')$ if $\delta>\delta'$. Suppose that $S=S(a_1, \cdots, a_k; 1/k)$ and $S'=S(a_1', \cdots, a_q'; 1/q)$ are in $\mathfrak{S}'$ and that $S\subset S'$. By the argument used in Theorem 3.3, each a_j' is a linear combination of the a_i, $i\leq k$; hence each a_j' must be an a_i, so $q\leq k$ and $f^{-1}(S)>f^{-1}(S')$.

This shows that Δ^ν is a cofinal part of $\mathfrak{S}$ if A is not fd; $\nu\neq 0$ since $\Delta^0>S$ if $\nu=0$ and no function on $\mathfrak{S}$ to U can exist satisfying Theorem 3.3. ν is unique since $\mathfrak{S}$ and Δ^ν are cofinally similar and (see §1) no set can be cofinally similar to two different stacks. If A is fd, $\mathfrak{S}\sim\mathfrak{N}\sim\Delta^0$, so no Δ^ν with $\nu>0$ can be a cofinal part of $\mathfrak{S}$ in this case.

If A is fd, let $\nu(A)=0$; if A is not fd, let $\nu(A)$ be the ordinal greater than 0 whose existence is asserted by this theorem.

COROLLARY 3.1. *If $\eta(A)=\min \nu(A')$ where the minimum is taken over all non-*fd *linear images A' of A, and if $X>\Delta^{\eta(A)}$, then $[A, X]\in\mathfrak{P}_{bd}$.*

Since the set of ordinals $\nu(A')$ is well-ordered by magnitude, there is a smallest one, so $\eta(A)$ is defined; let A' be an image of A for which $\nu(A')=\eta(A)$. By Theorems 3.3, 3.5, and 3.2, $[A', \Delta^{\nu(A')}]\in\mathfrak{P}_{bd}$, so, by 3.2, $[A, X]\in\mathfrak{P}_{bd}$ if $X>\Delta^{\eta(A)}$.

COROLLARY 3.2. *If $A=B^{(2n+1)}$, $[A, X]\in\mathfrak{P}_{bd}$ if $X>\Delta^{\nu(B^*)}$; if $A=B^{(2n+2)}$, $[A, X]\in\mathfrak{P}_{bd}$ if $X>\Delta^{\nu(B^{**})}$*[11].

This follows from Corollary 3.1 and this theorem: *Let A be isomorphic to a conjugate space and let A_1 be the subset of A^{**} consisting of all those points $\mathfrak{a}_a$ defined for each a in A by $\mathfrak{a}_a(\alpha)=\alpha(a)$ for every α in A^*; then there is a projection of A^{**} into A_1*[12].

If Y is any set of points y and B is any Banach space, there are certain easily defined Banach spaces of functions f on Y to B[13]. Let $m(Y, B)$ be the

---

[11] $B^{(n)}$ is defined by induction from $B^{(0)}=B$, $B^{(n+1)}=(B^{(n)})^*$.

[12] This need only be proved if $A=B^*$ for some B; in this case reducing each $\mathfrak{a}$ defined over B^{**} to a function defined only over B_1 defines a transformation of A^{**} into A: mapping back to A_1 by the usual method gives the desired projection. Phillips [23] has shown that c_0 is not the range of a projection of $m=c_0^{**}$; so some restriction on A is needed; it is not known whether A is isomorphic to a conjugate space if A_1 is the range of a projection in A^{**}.

[13] Most of the results given in this paragraph for these spaces of functions on Y to B can be adapted to the more general spaces of functions f on Y for which the value $f(y)$ always lies in some fixed space B_y; if all $B_y=B$, this reduces to the case discussed in the text. For example, see [9] for one case where Y is countable.

space of those f's such that $\|f\| = \sup_y \|f(y)\| < \infty$; for any p with $1 \leq p < \infty$ let $l_p(Y, B)$ be the space of those f's for which $\|f\| = (\sum_y \|f(y)\|^p)^{1/p} < \infty$ ([14]); let $c_0(Y, B)$ be the set of f's for which $\{y \mid |f(y)| > \epsilon\}$ is a finite set for every $\epsilon > 0$. In the special case $Y = \omega$, $B = B_0$, these spaces reduce to the well known spaces m, l_p, and c_0. It may be noted that the conjugate spaces of $c_0(Y, B)$, $l_1(Y, B)$, and $l_p(Y, B)$ with $1 < p < \infty$ are, respectively, equivalent to $l_1(Y, B^*)$, $m(Y, B^*)$ and $l_{p'}(Y, B^*)$ where $1/p + 1/p' = 1$.

If Y' is any subset of Y, let T be the operation which takes a function f on Y to B into the function Tf defined by $Tf(y) = f(y)$ if $y \in Y'$, $Tf(y) = \theta$ if $y \notin Y'$. Then it is clear that T defines a projection of norm 1 in each of the spaces $m(Y, B)$, $l_p(Y, B)$ and $c_0(Y, B)$ and that the range of T in these cases is equivalent to $m(Y', B)$, $l_p(Y', B)$ and $c_0(Y', B)$. Also, if Y is an infinite set, the spaces m, l_p, and c_0 are, respectively, linear images of $m[Y, B]$, $l_p[Y, B]$ and $c_0[Y, B]$.

COROLLARY 3.3 *If Y is any infinite set and B is any Banach space, $[m(Y, B), X]$, $[l_p(Y, B), X]$ and $[c_0(Y, B), X] \in \mathfrak{P}_{bd}$ if $X > \Delta^\gamma$ where $\aleph_\gamma$ is the power of the continuum.*

This follows from 3.2 and 3.5 since the power of a vector basis in m, l_p, or c_0, is that of the continuum.

The next result gives some conditions involving $\lambda(X)$; if A is not fd, let $\mu(A)$ be the smallest ordinal such that a fundamental set([15]) of power $\aleph_\mu$ exists in A.

THEOREM 3.6. *If $\lambda(X) > \mu(A) > 0$ and if $\lim_x U_x(a)$ exists for every a in A, then there is an x_0 such that $U_x = U_{x_0}$ if $x > x_0$, so $[A, X] \notin \mathfrak{P}_{bc}$. If $\lambda(X) > \nu(A)$, $[A, X] \notin \mathfrak{P}_{ac}$. If $\lambda(X) = \mu(A) > 0$, then $[A, X] \in \mathfrak{P}_{bc}$.*

Let A' be a fundamental set in A of power $\aleph_{\mu(A)}$; if $\lim_x U_x(a)$ exists for each a, then for each a in A' and integer k there is an x_{ak} in X such that $\|U_x(a) - U_{x'}(a)\| < 1/k$ if $x, x' > x_{ak}$. If $\lambda(X) > \mu(A)$, for each k there is an x_k which follows all x_{ak} so $\|U_x(a) - U_{x'}(a)\| < 1/k$ for all a in A' if $x, x' > x_k$. Since $\lambda(X) > 0$ there exists an x_0' following all x_k; so, if $x_0 > x_0'$, $U_x(a) = U_{x_0}(a)$ for all a in A' if $x > x_0$; hence $U_x(a) = U_{x_0}(a)$ for all a in A.

A similar argument if $\lambda(X) > \nu(A)$ and if $\lim\sup_x |\alpha_x(a)| < \infty$ for every a shows that there is an x_0 in X such that, for each a, a $k_a > 0$ exists with $|\alpha_x(a)| < k_a$ if $x > x_0$. That $\lim\sup_x \|\alpha_x\| < \infty$ follows from a theorem of Hildebrandt [15] which has as a special case this theorem([16]): *If A is a Banach*

([14]) If ϕ is a real-valued, non-negative function defined over Y, $\sum_y \phi(y)$ is defined to be $\sup_\eta \sum_{y \in \eta} \phi(y)$, where the supremum is taken over all finite subsets η of Y. Hence the assumption that $\sum_y \phi(y) < \infty$ implies that $\{y \mid \phi(y) > 0\}$ is at most countable.

([15]) A set A' is fundamental in A if the set of linear combinations of elements of A' is dense in A. It is known [18] that $\mu(A) = \nu(A)$ if and only if $\aleph_\mu$ is at least as great as the power of the continuum.

([16]) Banach's theorem is a corollary of Hildebrandt's in its more general form; the reader

space, if X is a directed set and if $\alpha_x \in A^*$, then $\lim\sup_x \|\alpha_x\| < \infty$ *if a sequence* $\{x_n\} \subset X$ *exists with the following property: For each a in A there exist integers* k_a *and* m_a *such that* $|\alpha_x(a)| < k_a$ *if* $x > x_{m_a}$.

If $\lambda(X) = \mu(A) > 0$, transfinite sequences $\{x_\nu\} \subset X$ and $\{a_\nu\} \subset A$ can be defined as follows: (1) ν ranges over all ordinals $< \omega_{\mu(A)}$, the first ordinal of power $\aleph_{\mu(A)}$; (2) $x_\nu > x_\rho$ if $\nu > \rho$ and $\{x_\nu\}$ has no upper bound; (3) the set $\{a_\nu \mid \nu < \omega_{\mu(A)}\}$ is fundamental in A. Let $n(\nu)$ be the largest integer such that $\nu - n$ is defined and define α_ν in A^* so that $\alpha_\nu(a_\rho) = 0$ if $\rho < \nu$ while $\|\alpha_\nu\| = n(\nu)$. If $\Omega = \{\nu \mid \nu < \omega_{\mu(A)}\}$, then $[A, \Omega] \in \mathfrak{P}_{bc}$, for if $a \in A$, there is a sequence $\{a_n'\}$ of linear combinations of the a_ν such that $\|a_n' - a\| \to 0$; hence there is a $\nu_a < \omega_{\mu(A)}$ such that a is in the smallest closed linear manifold containing $\{a_\nu \mid \nu < \nu_a\}$; hence $\alpha_\nu(a) = 0$ if $\nu > \nu_a$ so $\lim_\nu \alpha_\nu(a) = 0$ for every a in A; clearly $\lim\sup_\nu \|\alpha_\nu\| = \infty$. Also $X > \Omega$, for defining $g(\nu) = x_\nu$ and $h(x) = \inf \nu$ such that $x \not> x_\nu$ gives two functions with the desired properties. Therefore $[A, X] \in \mathfrak{P}_{bc}$ when $\lambda(X) = \mu(A) > 0$.

COROLLARY 3.4. *If $A = l_p(Y, B)$ or $c_0(Y, B)$, if Y is uncountable, and if $X > \Omega_1$, the set of denumerable ordinals, $[A, X] \in \mathfrak{P}_{bc}$. Under the hypothesis of the continuum $[m(Y, B), X] \in \mathfrak{P}_{bc}$ if $X > \Omega_1$ and Y is any infinite set.*

We conclude this section with a theorem giving some relationships among the sets $\mathfrak{P}$.

THEOREM 3.7. *If A exists for which $[A, X] \in \mathfrak{P}_d$, then $\lambda(X) = 0$. If $\lambda(X) = 0$, $[A, X] \in \mathfrak{P}_{ac}$ if and only if $[A, X] \in \mathfrak{P}_{bd}$, so $\mathfrak{P}_{ad} = \mathfrak{P}_{bd}$. $\mathfrak{P}_{bc} \neq \mathfrak{P}_{bd}$.*

If $[A, X] \in \mathfrak{P}_d$, there exist α_x in A^* and a sequence $\{x_n\} \subset X$ such that $\|\alpha_x\| > n$ if $x > x_n$; if $\lambda(X) > 0$, then an x_0 must exist following all x_n so that $\|\alpha_{x_0}\| = \infty$, which is impossible. If $\lambda(X) = 0$ and $[A, X] \in \mathfrak{P}_{ac}$, let $\{x_n\}$ be a monotone sequence with no upper bound in X and let α_x in A^* satisfy (a) and (c). Define h on X to X so that $h(x) > x$ for every x while $\|\alpha_{h(x)}\| > n^2$ if $x_n < x \not> x_{n+1}$; let $h(x) = x$ if $x \not> x_1$. Let $\alpha_x' = (1/n)\alpha_{h(x)}$ if $x_n < x \not> x_{n+1}$, $\alpha_x' = \alpha_x$ if $x \not> x_1$; then $\lim_x \alpha_x'(a) = 0$ for every a while $\lim_x \|\alpha_x\| = \infty$.

Since $\mathfrak{P}_{bd} \subset \mathfrak{P}_{ad} \subset \mathfrak{P}_{ac}\mathfrak{P}_d$, $\mathfrak{P}_{bd} = \mathfrak{P}_{ad}$. If X is the set of denumerable ordinals, or any other set such that $\lambda(X) > 0$, $[c_0(X, B_0), X] \in \mathfrak{P}_{bc}$ by the construction in Theorem 3.4; by the first statement of this theorem $[c_0(X, B_0), X] \notin \mathfrak{P}_{bd}$.

COROLLARY 3.5. *If $[A, Y] \in \mathfrak{P}_{ac}$, if $X > Y$, and if $X > \Delta^0$, then $[A, X] \in \mathfrak{P}_{bd}$.*

The major problem remaining here is to reduce the number of pairs whose class is unknown. Other problems are these: (1) Is $\mathfrak{P}_{bc} = \mathfrak{P}_{ac}$? (2) A corollary of Theorem 4.2 is that if A' is the range of a projection in A and if $[A', X]$ is in some class, then $[A, X]$ is in the same class; is this true if A' is any closed linear subset of A? (3) A special case of (2) is to decide whether or not

---

will note that the characterization sought in this section is not to involve the special choice of α_x; in that sense Hildebrandt's theorem is a useful tool but not a result of the desired type.

there is an X such that $[c_0, X]$ is in some class while $[m, X]$ is not. This might also be settled by the answer to (4). Is $[A^*, X]$ in one of these classes if $[A, X]$ is?

4. **Totally measurable functions and real operators.** In this section let Y be an abstract set, let Υ be a field([17]) of subsets of Y, and let B be a Banach space. If E is any set in Υ, let ϕ_E, the characteristic function of E, be 1 on E and 0 on its complement. A function f on Y with values in B is called *simple* if there exist a finite number of sets E_i in Υ and points b_i in B such that $f=\sum_{i\leq k}\phi_{E_i}b_i$. Let V be the space of all functions f on Y to B for which there exists a sequence $\{f_n\}$ of simple functions which converges uniformly to f; if $\|f\|=\sup_{y\subset Y}\|f(y)\|$, V is a Banach space. In the special case in which $B=B_0$, the space of real numbers, call the space V_r. If β is any element of B^* and $f\in V$, the function βf defined by $\beta f(y)=\beta(f(y))$ is in V_r and $\|\beta f\|\leq\|\beta\|\ \|f\|$; if $\phi\in V_r$ and $b\in B$, the function ϕb defined by $\phi b(y)=\phi(y)b$ is in V and $\|\phi b\|=\|\phi\|\ \|b\|$; moreover these functions of the form ϕb, with ϕ in V_r and b in B, form a fundamental set in V, since every $\phi_E b$ is of this form. Also if $\phi\in V_r$ there exist β in B^* and b in B such that $\beta\phi b=\phi$; in fact any choice such that $\beta(b)=1$ will do.

Let $\mathfrak{U}$ be $V:B$, the space of linear operators on V to B. Gowurin [13] has shown that each U in $\mathfrak{U}$ can be defined by means of a certain integral: For each $E\in\Upsilon$ define $\Phi(E)$ in $B:B$ by the relation $\Phi(E)b=U(\phi_E b)$ for every b in B; then (1) each $\Phi(E)$ is a linear operator on B to B, (2) Φ is *additive*; that is, $\Phi(E_1)+\Phi(E_2)=\Phi(E_1+E_2)$ if E_1 and E_2 are disjoint sets in Υ and (3) Φ is *limited*; that is,

$$W\Phi(Y) = \sup\left\|\sum_{i\leq k}\Phi(E_i)b_i\right\| < \infty$$

where the supremum is taken over all choices of b_i with $\|b_i\|\leq 1$ and all partitions of Y into a finite number of disjoint sets $E_1, \cdots, E_k$ in Υ([18]). On the other hand each Φ satisfying these three conditions defines a U in $\mathfrak{U}$ by means of the Gowurin integral: If $f=\sum_{i\leq k}\phi_{E_i}b_i$, let $\int fd\Phi=\sum_{i\leq k}\Phi(E_i)b_i$; then $\|\int fd\Phi\|\leq W\Phi(Y)\|f\|$ so $\|\int(f_n-f_m)d\Phi\|\to 0$ if $\|f_n-f_m\|\to 0$. If $f\in V$, let $\{f_n\}$ be a sequence of simple functions converging to f and let $\int fd\Phi=\lim_n\int f_n d\Phi$. If $U(f)=\int fd\Phi$, then $U\in\mathfrak{U}$, $\|U\|=W\Phi(Y)$, and Φ is derived from U by the relation $U(\phi_E b)=\Phi(E)b$ for every b in B.

Some of these set functions are of the form $\Phi(E)=\Psi(E)I$ where I is the

([17]) Υ is a field if finite sums of sets in Υ are in Υ, if $Y\in\Upsilon$ and if complements of sets in Υ are in Υ. The reader is asked to distinguish between Υ, used here, and T used later in this section.

([18]) It is easily verified that if Φ is limited and additive and if $W\Phi(E)=\sup\|\sum_{i\leq k}\Phi(E_i)b_i\|$, where the supremum is taken over all finite partitions of E into disjoint sets $E_1, \cdots, E_k$ in Υ, and sequences of points $b_1, \cdots, b_k$ of norm less than or equal to 1, then $W\Phi$ is a bounded, real-valued non-negative function on Υ such that $W\Phi(E_1)\leq W\Phi(E_1+E_2)\leq W\Phi(E_1)+W\Phi(E_2)$ for every pair of disjoint sets in Υ.

identity transformation in B and Ψ is a bounded, additive, real-valued function on Υ. For such a Φ write $U(f)=\int fd\Psi$ instead of $\int fd\Phi$; then $\|U\|=V\Psi(Y)$ $=\sup\sum_{i\leq k}|\Psi(E_i)|$ where the supremum is taken over all partitions of Y into a finite number of disjoint sets of Υ. All the elements of V_r^* are of the form $\mathrm{T}(\phi)=\int\phi d\Psi$ for some bounded, additive real-valued Ψ defined on Υ(19). The correspondence τ associating U in $V:B$ and $\tau U=\mathrm{T}$ in V_r^* if $\mathrm{T}(\phi)=\int\phi d\Psi$ for all ϕ in V_r and $U(f)=\int fd\Psi$ for all f in V is one-to-one and norm preserving between V_r^* and a subset $\mathfrak{U}_r$ of $V:B$. The reader can easily prove the following results:

(1) *If* $\beta\in B^*$, $U\in\mathfrak{U}_r$ *and* $f\in V$, *then* $\beta(U(f))=\tau U(\beta f)$.

(2) $U\in\mathfrak{U}_r$ *if and only if* $U(f)$ *is a linear combination of the values of* f *whenever* f *is a simple function in* V.

Since $\mathfrak{U}_r$ is a subset of $\mathfrak{U}$, the topologies defined in §2 impose three topologies in $\mathfrak{U}_r$. If $\mathfrak{N}$, $\mathfrak{S}$ and $\mathfrak{W}$ are the norm, s^* and w^* neighborhood systems of θ in $\mathfrak{U}$, let $\mathfrak{N}_r$, $\mathfrak{S}_r$ and $\mathfrak{W}_r$ be the intersection of these with $\mathfrak{U}_r$; that is, $N_r\in\mathfrak{N}_r$ if there is an N in $\mathfrak{N}$ such that $N_r=N\mathfrak{U}_r$, and similarly for $\mathfrak{S}_r$ and $\mathfrak{W}_r$.

THEOREM 4.1. $\mathfrak{N}_r\simeq\mathfrak{S}_r$ *if and only if* V_r *is* fd; $\mathfrak{S}_r\simeq\mathfrak{W}_r$ *if and only if one of the spaces* B *or* V_r *is* fd. *Hence* $\mathfrak{N}_r\simeq\mathfrak{W}_r$ *if and only if* V_r *is* fd.

If V_r is fd, there exist $\phi_1,\cdots,\phi_k$ in V_r which form a basis in V_r, so that for each $\epsilon>0$ there is a $\delta>0$ such that $\|\mathrm{T}\|<\epsilon$ if $|\mathrm{T}(\phi_i)|<\delta$ for $i=1,\cdots,k$. Take β_i in B^* and f_i in V, say $f_i=\phi_ib_i$, such that $\beta_if_i=\phi_i$ for $i=1,\cdots,k$. If U is in $W_r=W_r(f_1,\cdots,f_k,\beta_1,\cdots,\beta_k;\delta)$, then $|\beta_i(U(f_i))|<\delta$ for all i, but

$$|\beta_i(U(f_i))|=|\mathrm{T}(\beta_if_i)|=|\mathrm{T}(\phi_i)|$$

so $\|U\|=\|\mathrm{T}\|<\epsilon$ and $U\in N_r(\epsilon)$; that is, if V_r is fd, there is a W_r contained in each N_r, so $\mathfrak{N}_r\simeq\mathfrak{S}_r\simeq\mathfrak{W}_r$.

If B is fd, $\mathfrak{S}\simeq\mathfrak{W}$; so $\mathfrak{S}_r\simeq\mathfrak{W}_r$.

If V_r is not fd, consider two classes of neighborhoods $S_r(f_1,\cdots,f_k;\epsilon)$, first taking all f_i to be simple functions. Let $f_i=\sum_{j\leq k_i}\phi_{E_{ij}}b_{ij}$; then, for every β in B^*, $\beta f_i=\sum_{j\leq k_i}\phi_{E_{ij}}\beta(b_{ij})$, so the set of functions $\{\beta f_i \mid \beta \text{ in } B^*, i=1,\cdots,k\}$ can all be expressed as linear combinations of the characteristic functions of the finite number of sets E_{ij}. Hence the smallest linear manifold in V_r containing all the βf_i is fd and therefore does not fill up V_r; by a lemma of Banach [3, p. 57] there exists an T in V_r^* such that $\mathrm{T}(\beta f_i)=0$ for all f_i and β, while $\|\mathrm{T}\|>0$. Let $U=\tau^{-1}\mathrm{T}$, then $\beta(U(f_i))=\mathrm{T}(\beta f_i)=0$ so $U(f_i)=\theta$ for each i while $\|U\|=\|\mathrm{T}\|>0$. For each $K>0$ there is a point $U'=KU/\|U\|$ such that $\|U'\|=K$ and $U'\in S_r(f_1,\cdots,f_k;\epsilon)$.

This holds if the f_i are simple functions; if $f_1,\cdots,f_k$ are any functions in V, there exist k sequences of simple functions $\{f_{in}\}$ such that $\|f_{in}-f_i\|<2^{-n}$

(19) See Kantorovich and Fichtenholz [11], or specialize Gowurin's integral.

for each $i \leqq k$. Hence, for any $U \in \mathfrak{U}_r$, $\|U(f_{in}-f_i)\| < 2^{-n}\|U\|$. For any $K>0$ and $\epsilon>0$ take n_0 so that $2^{-n_0} < \epsilon/4K$; then in $S(f_{in_0}, \cdots, f_{kn_0}; \epsilon/4)$ there is a point U with $\|U\|=K$. $\|U(f_{in_0}-f_i)\| < K2^{-n_0} < \epsilon/4$ so $\|U(f_i)\| < \epsilon/2$ and $U \in S_r(f_1, \cdots, f_k; \epsilon)$ while $\|U\|=K$. This construction can be carried through for any S_r in $\mathfrak{S}_r$; so no S_r lies in an N_r if V_r is not fd.

If neither V_r nor B is fd, there is an f_0 in V such that no fd subspace of B contains all the values of f_0; consider $S_r(f_0; \epsilon)$ and any $W_r = W_r(f_1, \cdots, f_k; \beta_1, \cdots, \beta_k; \delta)$. Let $y_1, \cdots, y_{k+1}$ be points of Y for which the $k+1$ points $f_0(y_i)$ in B are linearly independent. Then $k+1$ numbers t_j exist which are not all zero and which satisfy the k equations $\sum_{j \leqq k+1} t_j \beta_i(f_i(y_j)) = 0$, $i=1, \cdots, k$. Let U in $\mathfrak{U}_r$ be defined by $U(f) = \sum_{j \leqq k+1} t_j f(y_j)$; then $U \in W_r$ since $\beta_i(U(f_i)) = \beta_i(\sum_{j \leqq k+1} t_j f_i(y_j)) = \sum_{j \leqq k+1} t_j \beta_i(f_i(y_j)) = 0$, while $\|U(f_0)\| > 0$ since the points $f_0(y_1), \cdots, f_0(y_{k+1})$ are linearly independent. For every K, $KU \in W_r$ but for K large enough $KU \notin S_r$; hence $\mathfrak{S}_r$ is not equivalent to $\mathfrak{W}_r$ in this case.

One remark on the problem of boundedness may be made for such spaces and operations.

COROLLARY 4.1. *For any B, V_r is a linear image of V; hence $\eta(V) \leqq \nu(V_r)$. If $X > \mathfrak{S}_r$ and V_r is not* fd, *U_x in $\mathfrak{U}_r$ can be chosen so that* $\lim_x \|U_x(f)\| = 0$ *for every f while* $\lim_x \|U_x\| = \infty$. *If V_r is* fd, *then no X and U_x in $\mathfrak{U}_r$ exist which satisfy these conditions.*

Note that V_r is fd if and only if Υ has only a finite number of distinct elements; V is fd if and only if both B and V_r are fd.

The next lemma is used in §6.

LEMMA 4.1. *If U_x and $U_0 \in \mathfrak{U}_r$ and if $\Upsilon_x = \tau U_x$, then $U_0 = w^* - \lim_x U_x$ if and only if $\Upsilon_0 = \tau U_0 = w^* - \lim_x \Upsilon_x$.*

$U_0 = w^* - \lim_x U_x$ if and only if $\lim_x \beta(U_x(f)) = \beta(U_0(f))$ for every β in B^* and f in V; that is, if and only if $\Upsilon_0(\beta f) = \lim_x \Upsilon_x(\beta f)$ for every β in B^* and f in V. But the set of such βf fills up V_r; so this is true if and only if $\Upsilon_0(\phi) = \lim_x \Upsilon_x(\phi)$ for every ϕ in V_r; that is, if and only if $\Upsilon_0 = w^* - \lim_x \Upsilon_x$[20].

5. **A completely additive integral.** That the class of totally measurable functions is rather limited is clear from the fact that the set of values of such a function f is a totally bounded subset of B; that is, for each $\epsilon>0$ there is a finite set of spheres of radius ϵ which together cover the set of values of f. If a measurable function is defined to be a function which is the limit of a pointwise convergent sequence of simple functions, then the class of bounded, measurable functions includes the class of totally measurable functions; the two classes are the same only if B is fd or if Υ has only a finite number of

[20] This lemma is a restatement of the fact that the equivalence τ of V_r^* and $\mathfrak{U}_r$ carries $\mathfrak{W}_r$ into the set of weak* neighborhoods of θ in V_r^*. Note that the proof of (2) of Theorem 6.2 shows that the w^* and s^* topologies are the same in the unit sphere of $\mathfrak{U}_r$ although they are different in the unit sphere of $\mathfrak{U}$ itself.

distinct elements. Birkhoff [4], Bochner [6], Dunford [10], Gelfand [12], Pettis [21], Phillips [22], and Price [24], among others, have defined and studied integrals of a Lebesgue-like nature for functions with values in a Banach space; Gowurin [13] and Bochner and Taylor [7] have considered a "Riemann-Stieltjes" integral; as far as I know, no attempt except in [24] has been made to define a completely additive integral similar to Gowurin's.

In this section Y is any set and B is a Banach space([21]); Υ is further restricted to be a σ-field([22]). A function f on Y to B will be called *half-simple* if it is bounded; that is, if there is a $K>0$ such that $\|f(y)\| \leqq K$ for all y, and if there exist a countable number of disjoint sets $\{E_i\}$ in Υ such that f has the constant value b_i on E_i.

LEMMA 5.1. *$\mathfrak{B}$, the class of bounded measurable functions on Y to B, is the class of all functions on Y to B which can be uniformly approximated by half-simple functions; hence $\mathfrak{B}$ is a Banach space if $\|f\| = \sup_y \|f(y)\|$.*

Every half-simple function is in $\mathfrak{B}$; so every f which can be approximated uniformly by half-simple functions $\{f_n\}$ is in $\mathfrak{B}$; the construction of a sequence of simple functions converging to f is as follows: Suppose that $f_n(y) = b_{in}$ if $y \in E_{in}$ and enumerate $\{E_{in}\}$ and $\{b_{in}\}$ as single sequences $\{E_j'\}$ and $\{b_j'\}$. For each j let E_{kj}'' be any enumeration of the disjoint sets obtained by intersecting all possible combinations of the E_i', $i \leqq j$, and their complements. Define f_j' by $f_j'(y) = \theta$ on $Y - \sum_{i \leqq j} E_i'$, $f_j'(y) = b_{kj}''$ on E_{kj}'' where b_{kj}'' is a b_i', $i \leqq j$, for which $\sup_{y \in E_{kj}''} \|b_i' - f(y)\|$ is a minimum; then the f_j' are simple functions and converge pointwise to f. If $f \in \mathfrak{B}$, there is a sequence $\{f_n\}$ of simple functions such that $\|f_n(y) - f(y)\| \to 0$ for each y in Y and $\|f_n\| \leqq \|f\|$; enumerate the values of these f_n in a sequence $\{b_i\}$. For any $\epsilon > 0$ let $E_i' = \{y \mid \|f(y) - b_i\| < \epsilon\}$; then $\sum_i E_i' = Y$ and each $E_i' \in \Upsilon$, because $y \in E_i'$ if and only if there is an n such that, for every $k > n$, $\|f_k(y) - b_i\| < \epsilon$; that is, if and only if $y \in \sum_n \prod_{k>n} E_{ki}$ where $E_{ki} = \{y \mid \|f_k(y) - b_i\| < \epsilon\}$. Each E_{ki} is in Υ, so each E_i' is also in Υ; let $E_1 = E_1'$, $E_{i+1} = E_{i+1}' - \sum_{k \leqq i} E_k'$, and define f_ϵ by $f_\epsilon(y) = b_i$ if $y \in E_i$. Since the E_i are disjoint, are in Υ, and cover Y, f_ϵ is half-simple; clearly $\|f - f_\epsilon\| < \epsilon$ so every element of $\mathfrak{B}$ can be approximated uniformly by half-simple functions.

From this lemma it is clear that the difficulties of defining an integral for functions in $\mathfrak{B}$ are mostly concentrated in defining the integral of every half-simple function. Gelfand [12] calls a series $\sum_i b_i$ of points of a Banach space B *unconditionally convergent* if $\sum_i |\beta(b_i)| < \infty$ for every β in B^*. If Δ^0 is the stack whose elements are the finite sets, δ, of positive integers, Alaoglu has shown

(21) Throughout this section we shall consider that B is imbedded in its second conjugate space B^{**} by the usual transformation associating b in B with $\mathfrak{b}_b$ in B^{**} if $\mathfrak{b}_b(\beta) = \beta(b)$ for every β in B^*.

(22) Υ is a σ-field (sometimes called a Borel field) if Υ is a field and if the sum of any countable collection of sets of Υ is a set of Υ.

that $\sum_i b_i$ converges unconditionally if and only if $\lim_\delta \sum_{i\in\delta} \beta(b_i)$ converges for every β in B^* and if and only if $\sup_\delta \|\sum_{i\in\delta} b_i\| < \infty$. Let the *sum* of the unconditionally convergent series $\sum_i b_i$ be that point $\mathfrak{b}$ of B^{**} for which $\mathfrak{b}(\beta) = \lim_\delta \beta(b_\delta)$ for every β in B^*, where $b_\delta = \sum_{i\in\delta} b_i$; then $\|\mathfrak{b}\| \leqq \lim \sup_\delta \|b_\delta\|$.

If Φ is a limited, additive function defined over the σ-field Υ with values in $B:B$, if $\{E_i\}$ is any sequence of disjoint sets in Υ, and if $\{b_i\}$ is any sequence of points of norm not exceeding 1, then $\sum_i \Phi(E_i)b_i$ is unconditionally convergent because $\|\sum_{i\in\delta} \Phi(E_i)b_i\| \leqq W\Phi(Y)$ for every δ. Φ is called completely additive (symbol: ca) if for each b in B and sequence $\{E_i\}$ of disjoint sets of Υ, $\sum_i \Phi(E_i)b = \Phi(\sum_i E_i)b$; complete additivity clearly implies finite additivity. A theorem of Orlicz [3, p. 240, (3)] asserts that if every subseries of a given series converges in Gelfand's sense to a point of B, then the series converges in Orlicz' sense; that is, $\lim_\delta \|\sum_{i\in\delta} b_i - \sum_i b_i\| = 0$. Hence, if Φ is ca,

$$\left\| \sum_{i\in\delta} \Phi(E_i)b - \Phi\left(\sum_i E_i\right)b \right\| \to 0$$

for every b in B and sequence $\{E_i\}$ of disjoint sets of Υ.

If Φ is ca and limited and if f is a half-simple function with the values b_i on the disjoint sets E_i of Υ, let $\int f d\Phi$ be the sum of the series $\sum_i \Phi(E_i)b_i$; from the complete additivity of Φ it is easily shown that this sum is independent of the decomposition $\{E_i\}$ as long as f is constant over each set E_i. The argument used to show that $\sum_i \Phi(E_i)b_i$ is unconditionally convergent also shows that $\|\int f d\Phi\| \leqq \|f\| W\Phi(Y)$ if f is half-simple. If f is any element of $\mathfrak{B}$, let $\{f_n\}$ be a sequence of half-simple functions converging uniformly to f; then the points $\mathfrak{b}_n = \int f_n d\Phi$ form a Cauchy sequence in B^{**} and must converge to some point of B^{**}; let $\int f d\Phi = \lim_n \int f_n d\Phi$. This value is easily shown to be independent of the choice of the sequence $\{f_n\}$ converging uniformly to f. If $U(f) = \int f d\Phi$, then U is a linear operator on V with values in B^{**} and $\|U\| = W\Phi(Y)$.

In many cases it is desirable to have $\int f d\Phi$ in B for every f in $\mathfrak{B}$. This is equivalent to requiring that $\int f d\Phi$ be in B for every half-simple function f; that is, to requiring that $\sum_i \Phi(E_i)b_i$ be in B for every sequence $\{E_i\}$ of disjoint sets of Υ and every bounded sequence $\{b_i\}$ of points of B. By the theorem of Orlicz mentioned before, this is equivalent to requiring that $\sum_i \Phi(E_i)b_i$ converge in Orlicz' sense for every such choice of $\{E_i\}$ and $\{b_i\}$; Φ will be called *convergent* if this last condition holds. From this we have the following result.

THEOREM 5.1. *If Φ is limited and* ca, *$\int f d\Phi \in B$ for each f in $\mathfrak{B}$ if and only if Φ is convergent; hence $\int f d\Phi \in B$ if B is weakly sequentially complete*([22]) *or is*

([22]) B is weakly sequentially complete if the existence of $\lim_n \beta(b_n)$ for every β in B^* implies that a b_0 in B exists such that $\lim_n \beta(b_n) = \beta(b_0)$ for every β in B^*.

reflexive or if Φ is of bounded variation[24] *or if $\Phi = \Psi T$ where Ψ is a real-valued,* ca *set-function and T is any element of $B:B$.*

THEOREM 5.2. *If Φ is additive and convergent, $W\Phi(Y) < \infty$; if Φ is* ca, *then $W\Phi(\sum_i E_i) = \lim_n W\Phi(\sum_{i<n} E_i)$; if Φ is* ca *and convergent, then $W\Phi(E_k)$ decreases to zero for every decreasing sequence $\{E_k\}$ of sets of Υ with empty intersection.*

Suppose that Φ is additive and convergent and that $W\Phi(Y) = \infty$; say that a set Y_0 of Υ has property (A) if Y_0 has two disjoint subsets Y_0' and Y_0'' such that $W\Phi(Y_0') = \infty$ and $W\Phi(Y_0'') = \infty$. Then only a finite number of disjoint sets of Υ can have property (A) for if an infinite number have this property, then there would exist a sequence $\{Y_i\}$ of disjoint sets of Υ such that $W\Phi(Y_i) = \infty$ for each i. Choice of E_{ij} in Υ and b_{ij} of norm not greater than one could then be made so that the series $\sum_{ij} \Phi(E_{ij})b_{ij}$ reordered in any way as a simple series would have unbounded partial sums, thus contradicting convergence of Φ.

Therefore there exist disjoint sets $Y_1, \cdots, Y_k$, $k \geqq 1$, whose sum is Y, such that $E \subset Y$, and $W\Phi(Y_i - E) = \infty$ imply that $W\Phi(E) < \infty$; let Y_0 represent any one of these Y_i. Define the sequence $\{E_i\} \subset \Upsilon$ of subsets of Y_0 as follows: If among the sets $E \subset Y_0$ such that $W\Phi(Y_0 - E) = \infty$ there are any such that $W\Phi(E) \neq 0$, let n be the smallest integer such that such an E_1 exists with $W\Phi(E_1) > 1/n$; if E_i, $i < k$, are defined and disjoint, let n_k be the smallest integer such that a subset E_k of $Y_0 - \sum_{i<k} E_i$ has $W\Phi(E_k) > 1/n_k$ and $W\Phi(Y_0 - E_k) = \infty$.

If an n_0 exists such that $W\Phi(E_i) > 1/n_0$ for an infinite sequence of these E_i, then a series $\sum_{ij} \Phi(E_{ij}')b_{ij}'$ could be found with partial sums not tending to zero which again contradicts convergence, so $W\Phi(E_i) \to 0$ as $i \to \infty$. Consider $E_0 = Y - \sum_i E_i$, and $\sum_i E_i$; by the definition of Y_0 either $W\Phi(E_0)$ or $W\Phi(\sum_i E_i) < \infty$. If $W\Phi(\sum_i E_i) < \infty$, then $W\Phi(E_0) = \infty$ since $W\Phi(Y_0) = \infty$; hence there exist a sequence $\{E_{kj}\}$ of partitions of E_0 and a sequence $\{b_{kj}\}$ of points of norm not exceeding 1 such that

$$\left\| \sum_{j \leqq n} \Phi(E_{kj})b_{kj} \right\| > k;$$

the partition $\{E_{k+1j}\}$ may be made a refinement of the partition $\{E_{kj}\}$. By definition of E_0, if $E \subset E_0$ and $W\Phi(E_0 - E) = \infty$, then $W\Phi(E) = 0$, since otherwise E would have been chosen among the E_i at some stage. Hence for each k there is a j_k such that $\|\Phi(E_{kj})\| = 0$ if $j \neq j_k$, so, letting $E_k' = E_{kj_k}$ and $b_k' = b_{kj_k}$, $\|\Phi(E_k')b_k'\| > k$ and $W\Phi(E_k') = \infty$. Therefore $E_{k+1}' \subset E_k'$ and $W\Phi(E_k' - E_{k+1}') = 0$, so $\Phi(E_{k+1}') = \Phi(E_k')$ for every k. It follows that $\|\Phi(E_i')b_k'\| \to \infty$ while

[24] Φ is of bounded variation if $\sup \sum_{i \leqq k} \|\Phi(E_i)\| < \infty$ where the supremum is taken over all partitions of Y into a finite number of disjoint sets of Υ.

$\|b_k'\| \leqq 1$, but this is impossible since $\Phi(E_1)$ is a linear operator on B, so $W\Phi(E_0) < \infty$.

This leaves the alternative hypothesis that $W\Phi(\sum_i E_i) = \infty$. A more careful repetition of the preceding argument, letting $\{E_{kj}\}$ be a partition of $\sum_{i>k} E_i$, leads to a contradiction here. This shows that $W\Phi(Y_0)$ must be finite, but Y_0 was any one set in a finite partition of Y so $W\Phi(Y) < \infty$ also.

Next assume that Φ is ca. Clearly the sets E_i mentioned can be taken to be disjoint; suppose then that

$$W\Phi\left(\sum_i E_i\right) > K = \lim_n W\Phi\left(\sum_{i \leqq n} E_i\right);$$

then there exists a partition $E^1, \cdots, E^k$ of $\sum_i E_i$ into disjoint sets of Υ and a set of points $b_1, \cdots, b_k$ of norm not greater than 1 such that

$$\left\| \sum_{j \leqq k} \Phi(E^j) b_j \right\| > K + 2\epsilon$$

for some $\epsilon > 0$. By complete additivity of Φ, n can be chosen so large that

$$\left\| \Phi(E^j) b_j - \sum_{i \leqq n} \Phi(E_i E^j) b_j \right\| < \frac{\epsilon}{2k}$$

for each $j \leqq k$, so $\|\sum_{j \leqq k} \sum_{i \leqq n} \Phi(E_i E^j) b_j\| > K + \epsilon$. The sets $E_i E^j$, $i \leqq n$, $j \leqq k$, form a partitition of $\sum_{i \leqq n} E_i$ so $K + \epsilon < W\Phi(\sum_{i \leqq n} E_i) \leqq K$; this contradiction shows that $W\Phi(\sum_i E_i) \leqq \lim_n W\Phi(\sum_{i \leqq n} E_i)$. Since $W\Phi(E)$ increases with E, the conclusion holds.

The other conclusion is a simple consequence of these two; the assumptions that $W\Phi(E_k) \downarrow 2\epsilon > 0$ and that Φ is ca show that there exists a sequence $\{k_i\}$ such that $W\Phi(E_{k_i} - E_{k_{i+1}}) > \epsilon$ for every i; this contradicts convergence.

If $B = B_0$, the space of real numbers, each $\Phi(E)$ is a real number and Φ is convergent if and only if $\sum_i \Phi(E_i) t_i$ converges for every bounded sequence $\{t_i\}$ of real numbers and every sequence $\{E_i\}$ of disjoint sets of Υ; that is, if and only if $\sum_i |\Phi(E_i)| < \infty$ for each such sequence $\{E_i\}$.

COROLLARY 5.1. *If and only if the real-valued, additive set-function Ψ on Υ has the property that $\sum_i \Psi(E_i)$ converges for every sequence $\{E_i\}$ of disjoint sets of Υ, $V\Psi(Y) < \infty$.*

This is true since $V\Psi(Y) = W\Psi(Y)$ in this case. Since a ca real-valued set-function has the property that

$$\sum_i |\Psi(E_i)| = \Psi(E') + |\Psi(E'')|$$

where E' is the sum of those E_i such that $\Psi(E_i) \geqq 0$, and E'' is the sum of those E_i such that $\Psi(E_i) < 0$; from this follows a well known theorem.

COROLLARY 5.2. *A* ca *real-valued set-function is of bounded variation.*

If B is any fd space and the values of Φ are in $B:B$, both these corollaries hold for such a Φ, although $V\Phi(Y)$ and $W\Psi(Y)$ are no longer so simply related.

THEOREM 5.3. *A linear operator U on V to B can be expressed in the form $U(f)=\int fd\Phi$ where Φ is* ca *and convergent if and only if $U(f)=\lim_n U(f_n)$ whenever $\|f_n\|$ is uniformly bounded and f_n converges pointwise to f.*

If f_n converges pointwise to f, if $E_{kn}=\{y \mid \|f(y)-f_m(y)\|>1/k \text{ if } m>n\}$, then for each k, $E_{kn}\downarrow 0$, so by the preceding lemma $\lim_n W\Phi(E_{kn})=0$ for each k. Then

$$\|U(f)-U(f_n)\| \leqq \left\|\int_{E_{kn}}(f-f_n)d\Phi\right\|+\left\|\int_{Y-E_{kn}}(f-f_n)d\Phi\right\|.$$

For given $\epsilon>0$ take $k>1/\epsilon$ and take n so large that $W\Phi(E_{kn})<\epsilon$; then

$$\|U(f)-U(f_n)\|<\epsilon\|f-f_n\|+\epsilon W\Phi(Y).$$

If U satisfies the last condition of the theorem, let $\Phi(E)$ be the operator on B to B defined by $\Phi(E)b=U(\phi_E b)$ for each b in B. Then Φ is additive and limited since $\|U\|=W\Phi(Y)$. If f is a simple function, $U(f)=\int fd\Phi$ is an element of B. If f is half-simple with values b_i on sets E_i,

$$\left\|\sum_{i\in\delta}\Phi(E_i)b_i\right\|=\left\|U\left(\sum_{i\in\delta}\phi_{E_i}b_i\right)\right\|\leqq\|f\|\,\|U\|,$$

but no matter in what order the sets E_i are arranged $\lim_n\sum_{i\leqq n}U(\phi_{E_i}b_i)=U(f)$ so $\sum_i \Phi(E_i)b_i$ is convergent in Orlicz' sense and Φ is convergent.

$$\sum_i\Phi(E_i)b=\lim_n\sum_{i\leqq n}U(\phi_E b_i)=\lim_n U\left(\sum_{i\leqq n}\phi_{E_i}b\right)=\lim_n U\left(\phi_{\sum_{i\leqq n}E_i}b\right)$$
$$=\Phi\left(\sum_i E_i\right)b$$

so Φ is ca.

In case Φ is equal to ΨI, where Ψ is real-valued, this integral is consistent with, say, Dunford's integral for bounded, measurable functions.

A desirable property for an integral is this: If f is in $\mathfrak{B}$ and T is in $B:B$, then $T\int fd\Phi=\int Tfd\Phi$, where Tf is the function in $\mathfrak{B}$ such that $Tf(y)=T(f(y))$. With this integral this does not hold for all T and f for two reasons, $\int fd\Phi$ may not lie in B, and T may not commute with all $\Phi(E)$. The first difficulty can be avoided; if T is an operator on B to B, define T^*, the adjoint([25]) of T, to

([25]) Banach [3, p. 99] calls T^* the conjugate of T and represents it by $\bar{T}$.

be that operator on B^* to B^* such that $T^*\beta(b)=\beta(Tb)$ for every b in B and β in B^*. Then (1) $\|T^*\|=\|T\|$, (2) $(T_1T_2)^*=T_2^*T_1^*$ and (3) if $T^{**}=(T^*)^*$, the operator T^{**} agrees with T over B.

THEOREM 5.4. *If Φ is limited and* ca, *T commutes with all $\Phi(E)$ if and only if $\int Tfd\Phi=T^{**}\int fd\Phi$ for every f in $\mathfrak{B}$.*

If there is an E in Υ such that $\Phi(E)$ does not commute with T, let b be a point such that $T\Phi(E)b\neq\Phi(E)Tb$ and let $f=\phi_E b$; then $T^{**}\int fd\Phi=T\Phi(E)b \neq\Phi(E)Tb=\int Tfd\Phi$.

If f is half-simple with values b_i on disjoint sets E_i, Tf has values Tb_i on the same sets E_i. $\int fd\Phi$ is that point $\mathfrak{b}$ of B^{**} for which

$$\mathfrak{b}(\beta) = \lim_{\mathfrak{s}} \sum_{i\in\mathfrak{s}} \beta(\Phi(E_i)b_i) = \lim_{\mathfrak{s}} \beta\left(\sum_{i\in\mathfrak{s}} \Phi(E_i)b_i\right).$$

$$\begin{aligned} T^{**}\mathfrak{b}(\beta) = \mathfrak{b}(T^*\beta) &= \lim_{\mathfrak{s}} T^*\beta\left(\sum_{i\in\mathfrak{s}} \Phi(E_i)b_i\right) = \lim_{\mathfrak{s}} \beta\left[T\left(\sum_{i\in\mathfrak{s}} \Phi(E_i)b_i\right)\right] \\ &= \lim_{\mathfrak{s}} \beta\left(\sum_{i\in\mathfrak{s}} T\Phi(E_i)b_i\right) = \lim_{\mathfrak{s}} \beta\left(\sum_{i\in\mathfrak{s}} \Phi(E_i)Tb_i\right) \\ &= \lim_{\mathfrak{s}} \sum_{i\in\mathfrak{s}} \beta(\Phi(E_i)Tb_i). \end{aligned}$$

But $\int Tfd\Phi$ is the point $\mathfrak{b}_1$ of B^{**} for which

$$\mathfrak{b}_1(\beta) = \lim_{\mathfrak{s}} \sum_{i\in\mathfrak{s}} \beta(\Phi(E_i)Tb_i) = T^{**}\mathfrak{b}(\beta)$$

for every β in B^*.

Since all the operators involved are continuous and since the half-simple functions are dense in $\mathfrak{B}$, the conclusion follows.

COROLLARY 5.3. *If Ψ is* ca *and real-valued, then $T\int fd\Psi=\int Tfd\Psi$ for every f in $\mathfrak{B}$ and every linear operator T on B to B.*

In this case $V\Psi(Y)<\infty$ so $\int fd\Psi\in B$ for each f in $\mathfrak{B}$. Every T in $B:B$ commutes with multiplication by real numbers.

This section closes with some examples. Let Y be the class of integers and let Υ be the class of all subsets of Y; for every real-valued function f on Y and every set E in Υ let $\Phi(E)f=\phi_E f$; that is, $\Phi(E)f(n)=f(n)$ if $n\in E$, $\Phi(E)f(n)=0$ if $n\notin E$. If B is $l_p=l_p(Y, B_0)$, $1\leqq p<\infty$, then $\Phi(E)f\in B$ if f does, so each $\Phi(E)\in B:B$ as $\|\Phi(E)\|\leqq 1$ for every E. This function is ca since $\sum_i \Phi(E_i)b=\Phi(\sum_i E_i)b$ for every b and sequence $\{E_i\}$ of disjoint sets in Y. However, $W\Phi(Y)=\infty$ so the theorem that a real-valued ca set-function is of bounded variation is not true if the words "real-valued" are deleted, even if "limited" replaces "of bounded variation."

If $B=c_0$ instead of l_p, each $\Phi(E)$ again defines a linear operator of norm less than or equal to 1 on B to B. Since $W\Phi(E)=1$ on every non-empty set E in Y, this gives an example of a function such that $W\Phi(Y)<\infty$ while Φ is not convergent. For example, defining b_i in B by $b_n(n)=1$, $b_i(n)=0$ if $i\neq n$, gives a sequence of points such that([26]) $\sum_n \Phi((n))b_n$ is in m instead of in c_0 since the sum is that f for which $f(n)=1$ for every n.

For one more example, take $B=m$ and let Φ be defined by $\Phi(E)b=\phi_E f_0 b$, where f_0 is the function for which $f_0(n)=1/n$. Then $\|\Phi(E)\|=W\Phi(E)=\sup_{n\in E} 1/n$, so Φ is convergent, but Φ is not of bounded variation since $\sum_n \|\Phi((n))\|=\sum_n 1/n=\infty$.

A more general integral is easily defined with properties almost precisely the same as those discussed here. If B and B' are two Banach spaces and if the set-function Φ has values in $B:B'$, then limited, ca, and convergent set-functions Φ can be defined almost as before; in this case $\int f d\Phi$ is a point in B'^{**} or, if Φ is convergent, in B'. An illustration is furnished by the last example above if the values of Φ are interpreted as transformations of m into c_0. Theorem 5.4 does not carry over to this case.

6. **General summability theorems.** Silverman and Toeplitz and others have given conditions on a matrix $\{a_{mn}\}$ of real numbers which are necessary and sufficient that it transform every convergent sequence $\{t_m\}$ into another convergent sequence $\{s_m\}$, where $s_m=\sum_n a_{mn}t_n$, which converges to the same limit. The theorem has a great many generalizations; one of these arises naturally from using functions on a directed set instead of sequences, another from letting the values of these functions be points of a Banach space instead of real numbers. The form of the theorem to be stated is suggested by the fact that c, the space of convergent sequences of real numbers, is a Banach space if $\|\{t_n\}\|=\sup_n |t_n|$; in fact it is a space of the form considered in §4 if the field is the smallest field containing all the finite sets of integers.

Let Y be any directed set and B any Banach space; let A be a Banach space whose elements are functions f on Y to B with the property that $\lim_y f(y)$ exists (in the norm topology) for each A. Define the operator L on A to B by setting $L(f)=\lim_y f(y)$ for each f in A; then L is additive and homogeneous but need not be continuous([27]). A set $A'\subset A$ is *dense in limit* in A if for each f in A and $\epsilon>0$ there is an f' in A' such that $\|f-f'\|<\epsilon$ and $\|L(f)-L(f')\|<\epsilon$.

Note that in the simple case $A=c$, above, the set A' of sequences which are ultimately constant is dense in c, and hence dense in limit in c because L is continuous in this case. The conditions (a) and (b) of Silverman-Toeplitz

---

([26]) (n) is the set whose only element is n.

([27]) The referee quite justly remarks that any additive homogeneous function L' on any Banach space A to B could be considered with similar results; for example, letting $L'(f)$ be the weak limit instead of the norm limit of f would give analogous results.

assure that every ultimately constant sequence will be taken into a sequence with the same limit.

If X is a directed set, for each x in X let U_x be a linear operator on A to B; the transformation $\{U_x\}$ thus defined on A to a class of functions on X to B is called *regular* on a subset A' of A if $\lim_x U_x(f)=L(f)$ for every f in A'. Clearly if $\{U_x\}$ is regular on A' and $A'\subset A''$, $\{U_x\}$ is regular on A''.

THEOREM 6.1. (1) *$\{U_x\}$ is regular on A if it satisfies the conditions* (a') *there exists a set A' dense in limit in A such that $\{U_x\}$ is regular on A', and* (b') $\lim\sup_x \|U_x\|<\infty$. (2) *If L is discontinuous on A and $\{U_x\}$ is regular on A,* $\lim\sup_x \|U_x\|=\infty$. (3) *If A is* fd, *if $\Delta^0>X$, if $\lambda(X)>\mu(A)$, or if for any other reason $[A, X]$ is not in $\mathfrak{P}_{bc}$, and if $\{U_x\}$ is regular on A, then* $\lim\sup \|U_x\|<\infty$ *and L is continuous.* (4) *If L is continuous on A, $\{U_x\}$ is regular on A if and only if $L=s^*-\lim_x U_x$.* (5) *If L is continuous and $X>\Delta^{\eta(A)}$ or if for any other reason $[A, X]\in\mathfrak{P}_{bc}$, there exists a $\{U_x\}$ regular on A such that* $\lim\sup_x\|U_x\|=\infty$ *(in the first case $\{U_x\}$ exists such that* $\lim_x\|U_x\|=\infty$).

(1) is a minor adaptation of a standard theorem on convergence of linear operators [3, p. 79, Theorem 3]. (2) is obvious since $\|L\|\leqq\lim\sup_x\|U_x\|$. (3) follows from various results of §3. (4) is a restatement of regularity on A. (5) follows from the definition of $\mathfrak{P}_{bc}$ and, for the last part, from Corollary 3.1 and Theorem 3.2.

In the special case in which A is a space V, as considered in §4, $\lim_y f(y)$ can exist for a simple function if and only if f is ultimately constant; in particular $\lim_y \phi_E b(y)$ exists if and only if either E or $Y-E$ is not cofinal in Y. The properties of cofinality mentioned after the definition show that if Υ_0 is a field of subsets of Y, and if Υ is the subclass of those sets E of Υ_0 for which either E or $Y-E$ is not cofinal in Y, then Υ is also a field. Use subscripts to indicate the field involved.

LEMMA 6.1. *If f can be uniformly approximated by functions simple Υ_0, and if $\lim_y f(y)$ exists, then f can be uniformly approximated by functions simple Υ.*

If $\epsilon>0$ is given, there exists a function $f_\epsilon=\sum_{i\leqq k}\phi_{E_i}b_i$ where the $E_i\in\Upsilon_0$ such that $\|f_\epsilon(y)-f(y)\|<\epsilon/3$ for all y in Y. Also there is a y_ϵ in Y such that $\|f(y)-b_0\|<\epsilon/3$ if $y>y_\epsilon$, where $b_0=\lim_y f(y)$. Let $E'=\sum E_i$ where the sum is taken over those E_i which contain a successor of y_ϵ. Define f_ϵ' on Y to B by $f_\epsilon'(y)=b_0$ if $y\in E'$, $f_\epsilon'(y)=f_\epsilon(y)$ if $y\notin E'$. Then $Y-E'$ is not cofinal in Y so $E'\in\Upsilon$; no E_i disjoint from E' can be cofinal in Y so the other E are also in Υ. Hence f_ϵ' is simple Υ, but $\|f_\epsilon'-f\|<\epsilon$.

From this lemma it follows that for this section it suffices to assume that Υ is a field of this special sort; that is, Υ satisfies (C): for each E in Υ either E or its complement is not cofinal in Y. In this case L is continuous on V, in fact $\|L\|\leqq 1$. It is clear from the criterion (2) of §4 that $L\in\mathfrak{U}_r$. Since the simple functions are dense in V, the condition (a') of (1), Theorem 6.1, for this

special space can be replaced by (a″) $\lim_x U_x(\phi_Y b)=b$ for each b in B; $\lim_x U_x(\phi_E b)=\theta$ for each b in B and each E in Υ such that E is not cofinal in Y. (2) of that theorem can not occur in this case; it is known that $\eta(V)\leqq\nu(V_r)$.

The special case in which A is a space V of this type while each U_x is in $\mathfrak{U}_r$ presents a situation more general than one studied by Vulich [28]. He considers convergent sequences $\{b_n\}$ of points of a Banach space and transformations U_m defined by means of a matrix of real numbers $\{a_{mn}\}$ so that $U_m(\{b_n\})=\sum_n a_{mn}b_n$ and this series converges absolutely for each $\{b_n\}$ so that $\sum_n |a_{mn}|<\infty$ for each m. Vulich proves that $\lim_m U_m(\{b_n\})=\lim_n b_n$ for every convergent sequence $\{b_n\}$ of points of B, if and only if the matrix satisfies the Toeplitz conditions; that is, if and only if the matrix defines a transformation regular on real sequences.

From Lemma 4.1 we have, letting $\Upsilon=\tau U$, as in §4.

LEMMA 6.2. *If each* $U_x\in\mathfrak{U}_r$, $\{U_x\}$ *is regular on* V *if and only if* $L=s^*-\lim_x U_x$; $\{\Upsilon_x\}$ *is regular on* V_r *if and only if* $L=w^*-\lim_x U_x$.

We use this to derive the following extension of Vulich's theorem.

THEOREM 6.2. *Let* Y *be a directed set,* Υ *a field of subsets of* Y *satisfying* (C), *and* V *the Banach space of functions totally measurable with respect to this field; Let* X *be a directed set, for each* x *in* X *let* U_x *be in* $\mathfrak{U}_r$, *and let* $\Upsilon_x=\tau U_x$. (1) *If* $\{U_x\}$ *is regular on* V, $\{\Upsilon_x\}$ *is regular on* V_r. (2) *If* $\{\Upsilon_x\}$ *is regular on* V_r *and* $\limsup_x \|U_x\|=\limsup_x\|\Upsilon_x\|<\infty$, *then* $\{U_x\}$ *is regular on* V. (3) *If* B *is* fd *and* $\{\Upsilon_x\}$ *is regular on* V_r, *then* $\{U_x\}$ *is regular on* V. (4) *If* V_r *is* fd, *or if* $\Delta^0>X$, *or if* $\lambda(X)>\mu(V_r)$, *or if for any other reason* $[V_r, X]\notin\mathfrak{P}_{bc}$, *then* $\limsup_x\|U_x\|<\infty$ *if* $\{\Upsilon_x\}$ *is regular on* V_r, *so, by* (2), $\{U_x\}$ *is regular on* V. (5) *If neither* V_r *nor* B *is* fd *and if* $X>\Delta^{\nu(V_r)}$, *then* U_x *can be chosen from* $\mathfrak{U}_r$ *so that* $\{\Upsilon_x\}$ *is regular on* V_r *while* $\{U_x\}$ *is not regular on* V.

(1) By (2) of §4, $L\in\mathfrak{U}_r$; if $\Lambda=\tau L$ and $L=s^*-\lim_x U_x=x^*-\lim_x U_x$, then $\Lambda=w^*-\lim_x \Upsilon_x$; clearly $\Lambda(\phi)=\lim_y\phi(y)$ for each ϕ in V_r.

(2) If $\phi\in V_r$ and $b\in B$, then $\phi b\in V$ and

$$\begin{aligned}\|U_x(\phi b)-L(\phi b)\| &= \sup_{\|\beta\|\leq 1}\left|\beta(U_x(\phi b))-\beta(L(\phi b))\right|\\ &= \sup_{\|\beta\|\leq 1}\left|\Upsilon_x(\beta(b)\phi)-\Lambda(\beta(b)\phi)\right|\\ &= \left|\Upsilon_x(\phi)-\Lambda(\phi)\right|\sup_{\|\beta\|\leq 1}\left|\beta(b)\right| = \|b\|\left|\Upsilon_x(\phi)-\Lambda(\phi)\right|.\end{aligned}$$

Hence $\|U_x(\phi b)-L(\phi b)\|\to 0$ for every ϕ in V_r and b in B if $\{\Upsilon_x\}$ is regular on V_r. But the set of all ϕb is fundamental in V, so, by (1) of Theorem 6.1, $\{U_x\}$ is regular on V.

(3) is true by Lemma 6.2 and Theorem 4.1. (4) follows from Theorems 4.1, 3.4 (the known half) and 3.6.

(5) If neither V_r nor B is fd, s^* and w^* topologies in $\mathfrak{U}_r$ are different; hence there is a neighborhood S_r of L which contains no w^* neighborhood of L. For each w^* neighborhood W of L let U_W in $\mathfrak{U}_r$ be in $W_r - S_r$; directing $\mathfrak{W}_r$ by inclusion gives $L = w^* - \lim_W U_W$ while L can not be $s^* - \lim_W U_W$. Since the weak* neighborhood system in V_r^* is isomorphic to $\mathfrak{W}_r$ as a directed set, $\Delta^{\nu(V_r)} \sim \mathfrak{W}_r$; in the usual manner if $X > \Delta^{\nu(V_r)}$, $\{U_x\}$ can be defined in terms of the $\{U_W\}$ to have the same properties.

As an example let us consider multiple sequences. Let Y be the set of n-tuples $y = y_1, \cdots, y_n$ of positive integers, directed by $y > y'$ if $y_i \geqq y_i'$, $i = 1, \cdots, n$, $n > 1$. Let B be any Banach space and let $bc(Y, B)$ be the set of those bounded functions f on Y to B for which $\lim_y f(y)$ exists. Then $bc(Y, B)$ is a subclass of $m(Y, B)$, which, since Y is countable, is a space $\mathfrak{B}$ of the sort considered in §5. Let X be the set of m-tuples $x = x_1, \cdots, x_n$ of positive integers, directed as Y is, and for each x in X let Φ_x be a convergent, ca set-function defined over all subsets of Y. Then $\Delta^0 > X$ and each Φ_x defines a linear operator U_x on $bc(Y, B)$ to B by the relation $U_x(f) = \int f d\Phi_x$ using the integral of §5.

THEOREM 6.3. *Under these conditions* $\{U_x\}$ *is regular on* $bc(Y, B)$ *if and only if* (1) $s^* - \lim_x \Phi_x(Y) = I$, (2) $\|U_x(f\phi_E)\| \to 0$ *if* $f \in bc(Y, B)$ *and if* E *is any set not cofinal in* Y, *and* (3) $\limsup_x W\Phi_x(Y) < \infty$.

If B is fd, (2) can be replaced by (2′) $s^* - \lim_x \Phi_x(E) = \theta$ (or $\lim_x \|\Phi_x(E)\| = 0$) if E is not cofinal in Y.

A smaller class of functions on this Y is the class $rc(Y, B)$; let Y' be a subset of Y which is directed by the same order relation holding in Y itself; then $f \in rc(Y, B)$ if and only if $\lim_{y'} f(y')$ exists for every such directed subset Y' of Y. A simple investigation shows that any such Y' has the following characteristics:

(a) There exists a set of integers $i_1, \cdots, i_j, \cdots, i_p$, $0 \leqq p \leqq n$, $i_j \leqq n$ for all $j = 1, \cdots, p$ and a set of integers $n_1, \cdots, n_p$, such that $y_{i_j} \leqq n_j$ for every $j = 1, \cdots, p$.

(b) The set Y'' of those y in Y' such that $y_{i_j} = n_j$ for $j = 1, \cdots, p$ is cofinal in Y' while $Y' - Y''$ is not.

Define a *slice* Y' of Y to be any set of the form $\{y \mid y_{i_j} = n_j \text{ for } j = 1, \cdots, p\}$ for any choice of $0 \leqq p \leqq n$, and $i_j \leqq n$: for example, if $n = 3$, the slices obtained by fixing 3, 2, 1 and 0 elements of each y are, respectively, single elements, columns, layers, and all of Y. Then each slice is a directed subset of Y (in case n elements of y are fixed we have the trivial directed set with one element) and the characteristics (a) and (b) show that Y' is a directed subset of Y if and only if there is a slice Y'' such that the intersection $Y''Y'$ is cofinal in both Y'' and Y' while $Y' - Y''$ is not cofinal in Y'. Letting Υ be the smallest field containing all the slices in Y, it is easily seen that $rc(Y, B)$ is the space V associated with this Υ.

Taking X as before to be the set of m-tuples $x_1, \cdots, x_m$, let Φ_x be any limited additive set-function on Y to $B:B$ and define $U_x(f) = \int f d\Phi_x$ for each f in $rc(Y, B)$, using the Gowurin integral.

THEOREM 6.4. *Under these last conditions* $\{U_x\}$ *is regular on* $rc(Y, B)$ *if and only if* (1) $s^*-\lim_x \Phi_x(Y) = I$, (2) $s^*-\lim_x \Phi_x(E) = 0$ *for every slice E not cofinal in* Y, *and* (3) $\lim\sup_x W\Phi_x(Y) < \infty$.

The usual modification if Φ is real-valued can be made.

These examples suffice to show something of the generality of the theorems of this section; Theorem 6.1 contains as special cases a number of theorems due to Toeplitz, Hamilton [14], Hill [16], the writer [8] and others. Its use is restricted by the requirement that the class of functions under discussion is a Banach space under some norm adapted to the problem; this is not the case, for example, of the class of all convergent double sequences(28). Further information about the problem of boundedness would also improve the results here.

BIBLIOGRAPHY

1. L. Alaoglu, *Weak topologies in normed linear spaces*, Annals of Mathematics, (2), vol. 41 (1940), pp. 252–267.

2. P. Alexandroff and H. Hopf, *Topologie* I, Berlin, 1935.

3. S. Banach, *Théorie des Opérations Linéares*, Warsaw, 1932.

4. G. Birkhoff, *Integration of functions with values in a Banach space*, these Transactions, vol. 38 (1935), pp. 357–378.

5. ———, *Moore-Smith convergence in general topology*, Annals of Mathematics, (2), vol. 38 (1937), pp. 39–56.

6. S. Bochner, *Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind*, Fundamenta Mathematicae, vol. 20 (1933), pp. 262–276.

7. S. Bochner and A. E. Taylor, *Linear functionals on certain spaces of abstractly-valued functions*, Annals of Mathematics, (2), vol. 39 (1938), pp. 913–944.

8. M. M. Day, *Regularity of function-to-function transformations*, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 296–303.

9. ———, *Reflexive Banach spaces not isomorphic to uniformly convex spaces*, Bulletin of the American Mathematical Society, vol. 47 (1941), pp. 313–317.

10. N. Dunford, *Integration in general analysis*, these Transactions, vol. 37 (1935), pp. 441–453.

11. G. Fichtenholz and L. Kantorovich, *Sur les opérations dans l'espace des functions bornées*, Studia Mathematica, vol. 5 (1934), pp. 69–98.

12. I. Gelfand, *Abstrakte Funktionen und lineare Operatoren*, Recueil Mathématique, (n.s.), vol. 4 (1938), pp. 235–284.

13. M. Gowurin, *Über die Stieltjiesche Integration abstrakter Funktionen*, Fundamenta Mathematicae, vol. 27 (1936), pp. 254–268.

14. H. J. Hamilton, *Transformations of multiple sequences*, Duke Mathematical Journal, vol. 2 (1936), pp. 29–60.

15. T. H. Hildebrandt, *On uniform limitedness of sets of functional operations*, Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 303–315.

(28) Hill and Hamilton [17] have discussed devices for avoiding this difficulty in the study of real-valued multiple sequences.

16. J. D. Hill, *A theorem in the theory of summability*, Bulletin of the American Mathematical Society, vol. 42 (1936), pp. 225–228.

17. J. D. Hill and H. J. Hamilton, *Operation theory and multiple sequence transformations*, Duke Mathematical Journal, vol. 8 (1941), pp. 154–162.

18. H. Löwig, *Über die Dimension linearen Räume*, Studia Mathematica, vol. 5 (1934), pp. 18–23.

19. E. H. Moore and H. L. Smith, *A general theory of limits*, American Journal of Mathematics, vol. 44 (1922), pp. 101–121.

20. J. von Neumann, *Zur Algebra der Funktionaloperatioren und Theorie der normalen Operatoren*, Mathematische Annalen, vol. 102 (1929–1930), pp. 913–944.

21. B. J. Pettis, *On integration in vector spaces*, these Transactions, vol. 44 (1938), pp. 277–304.

22. R. S. Phillips, *Integration in a convex linear topological space*, these Transactions, vol. 47 (1940), pp. 114–145.

23. ———, *On linear transformations*, these Transactions, vol. 48 (1940), pp. 516–541.

24. G. B. Price, *The theory of integration*, these Transactions, vol. 47 (1940), pp. 1–50.

25. A. E. Taylor, *The weak topologies of Banach spaces*, Proceedings of the National Academy of Sciences, vol. 25 (1939), pp. 438–440.

26. O. Toeplitz, *Über allgemeine lineare Mittelbildungen*, Prace Matematycznofizyczne, vol. 22 (1911), pp. 113–119.

27. J. W. Tukey, *Convergence and Uniformity in Topology*, Princeton, 1940.

28. B. Vulich, *Sur les methodes linéares de sommation dans les espaces abstraits*, Communications of the Institute of Mathematical and Mechanical Sciences of the University of Kharkhov and of the Mathematical Society of Kharkhov, (4), vol. 15 (1938).

29. J. V. Wehausen, *Transformations in linear topological spaces*, Duke Mathematical Journal, vol. 4 (1938), pp. 157–168.

University of Illinois,
Urbana, Ill.

THE RAMANUJAN IDENTITIES UNDER MODULAR SUBSTITUTIONS

BY

HANS RADEMACHER

1. **Introduction.** In connection with his discovery of certain divisibility properties of the partition function Ramanujan [1](1) stated the identities

$$\sum_{l=0}^{\infty} p(5l+4)x^l = 5\,\frac{\prod(1-x^{5m})^5}{\prod(1-x^m)^6}, \tag{1.1}$$

and

$$\sum_{l=0}^{\infty} p(7l+5)x^l = 7\,\frac{\prod(1-x^{7m})^3}{\prod(1-x^m)^4} + 49x\,\frac{\prod(1-x^{7m})^7}{\prod(1-x^m)^8}. \tag{1.2}$$

Here, as always in the sequel, the index m in the infinite products runs through all positive integers. If these identities, for which various proofs have been given, are expressed in terms of the Dedekind η-function

$$\eta(\tau) = e^{\pi i\tau/12}\prod(1-e^{2\pi i m\tau}), \qquad \Im(\tau) > 0, \tag{1.3}$$

they appear in a form which suggests certain group-theoretical considerations, similar to those employed by Hecke in his theory of modular forms. In this way we transform the identities into new ones which are noteworthy because of the occurrence of the Legendre symbol and which, by a simple further argument, lead also to a proof of (1.1) and (1.2). An analogous identity for the modulus 13, given by Zuckerman, can be treated in the same way.

G. N. Watson and H. S. Zuckerman have also derived identities for the moduli 5^2 and 7^2. These will lead us to certain modular equations, which in turn will shed some light on those identities.

PART I. IDENTITIES OF RAMANUJAN AND ZUCKERMAN

2. We have known since Euler

$$\sum_{n=0}^{\infty} p(n)x^n = \frac{1}{\prod(1-x^m)} \tag{2.1}$$

with $p(0)=1$, or

$$\frac{1}{\eta(\tau)} = e^{-\pi i\tau/12}\sum_{n=0}^{\infty} p(n)e^{2\pi i n\tau}. \tag{2.2}$$

Presented to the Society, February 22, 1941; received by the editors April 25, 1941.

(1) Numbers in square brackets refer to the bibliography at the end of this paper.

Hence we obtain

$$\sum_{\lambda=0}^{4} \frac{1}{\eta\left(\dfrac{\tau+24\lambda}{5}\right)} = \sum_{\lambda=0}^{4} e^{-(\pi i\tau/60)-(2\pi i\lambda/5)} \sum_{n=0}^{\infty} p(n) e^{(2\pi i n\tau/5)+(48\pi i\lambda n/5)}$$

$$= e^{-\pi i\tau/60} \sum_{n=0}^{\infty} p(n) e^{2\pi i n\tau/5} \sum_{\lambda=0}^{4} e^{-(2\pi i/5)\lambda(1-24n)}$$

$$= 5e^{-\pi i\tau/60} \sum_{n\equiv 4 \pmod 5} p(n) e^{2\pi i n\tau/5},$$

and therefore

$$\sum_{l=0}^{\infty} p(5l+4) e^{2\pi i l\tau} = \frac{1}{5} e^{-19\pi i\tau/12} \sum_{\lambda=0}^{4} \frac{1}{\eta\left(\dfrac{\tau+24\lambda}{5}\right)}.$$

If we also express the right-hand member of (1.1) in terms of $\eta(\tau)$, through its definition (1.3), we get

$$\sum_{\lambda=0}^{4} \eta\left(\frac{\tau+24\lambda}{5}\right)^{-1} = 5^2 \frac{\eta(5\tau)^6}{\eta(\tau)^6} \tag{2.3}$$

as a restatement of (1.1).

In a similar way (1.2) can be rewritten as

$$\sum_{\lambda=0}^{6} \eta\left(\frac{\tau+24\lambda}{7}\right)^{-1} = 7^2 \frac{\eta(7\tau)^3}{\eta(\tau)^4} + 7^3 \frac{\eta(7\tau)^7}{\eta(\tau)^8}. \tag{2.4}$$

3. We are now going to subject (2.3) and (2.4) to modular transformations, of which we need to test only the generators

$$S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The definition (1.3) shows that

$$\eta(\tau+24) = \eta(\tau).$$

Consequently the range of the summation in the left-hand member of (2.3) can be replaced by "modulo 5." Therefore S^{24} produces only a cyclical exchange of the terms of the sum and does not change the sum as a whole. It follows that S and S^{25} have the same effect on the left-hand member of (2.3), and since S^{25} means the replacement of $(\tau+24\lambda)/5$ by $((\tau+24\lambda)/5)+5$, this effect is clearly the appearance of a multiplier $e^{-5\pi i/12}$ in each summand. On the other side, the substitution S, that is, $\tau\rightarrow\tau+1$, provides the multiplier

$$e^{19\pi i/12} = e^{-5\pi i/12}$$

also on the right-hand side. Thus the equation (2.3) goes over into itself under the substitution S.

Similarly we see that S and S^{40} have the same effect on the left-hand member of (2.4), viz., multiplication of each summand by

$$e^{-7\pi i/12}.$$

This same factor is taken up also, as (1.3) shows, by each term of the right-hand member of (2.4) so that the equation (2.4) also remains invariant under the substitution S.

4. This is not so with T. Through the substitution

$$\tau \to -\tau^{-1}$$

equation (2.3) goes over into

$$\sum_{\lambda=0}^{4} \eta\left(\frac{-1+24\lambda\tau}{5\tau}\right)^{-1} = 5^2 \frac{\eta\left(-\dfrac{5}{\tau}\right)^5}{\eta\left(-\dfrac{1}{\tau}\right)^6}. \tag{4.1}$$

The left-hand member, which we designate by L_5, can be rewritten as

$$L_5 = \eta\left(\frac{-1}{5\tau}\right)^{-1} + \sum_{\lambda=1}^{4} \eta\left(\frac{24\lambda\,\dfrac{\tau+24\lambda'}{5} + b_\lambda}{5\,\dfrac{\tau+24\lambda'}{5} - 24\lambda'}\right)^{-1} \tag{4.2}$$

with

$$\lambda\lambda' \equiv -1 \pmod 5, \tag{4.21}$$

and

$$b_\lambda = -(24^2\lambda\lambda' + 1)/5. \tag{4.22}$$

Now for a modular substitution

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad c > 0,$$

we have [8],

$$\eta\left(\frac{a\tau+b}{c\tau+d}\right) = \exp\left(-\pi i\left(s(a,c) - \frac{a+d}{12c}\right)\right)(-i(c\tau+d))^{1/2}\eta(\tau), \tag{4.3}$$

where $s(a, c)$ is the "Dedekind sum"

$$s(a,c) = \sum_{\mu \bmod c} \left(\left(\frac{\mu}{c}\right)\right)\left(\left(\frac{a\mu}{c}\right)\right) \tag{4.31}$$

with

$$((x)) = \begin{cases} x - [x] - \frac{1}{2} & x \text{ not an integer,} \\ 0 & x \text{ an integer.} \end{cases} \tag{4.32}$$

If we apply (4.3) on (4.2) and the right-hand member of (4.1) for the modular substitutions

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

and

$$\begin{pmatrix} 24\lambda & b_\lambda \\ 5 & -24\lambda' \end{pmatrix},$$

we obtain after a few reductions

$$\frac{1}{5^{1/2}\eta(5\tau)} + \sum_{\lambda=1}^{4} \frac{\exp\left(-\pi i(s(\lambda, 5) + \frac{2}{5}(\lambda - \lambda'))\right)}{\eta\left(\dfrac{\tau + 24\lambda'}{5}\right)} = \frac{1}{5^{1/2}} \frac{\eta\left(\dfrac{\tau}{5}\right)^5}{\eta(\tau)^6}. \tag{4.4}$$

The Dedekind sums enjoy the following properties for

$$(h, k) = 1, \; hh' \equiv -1 \pmod{k}:$$

$$12ks(h, k) \equiv h - h' \pmod{k}, \tag{4.51}$$

$$12ks(h, k) \equiv 0 \pmod{3}, \qquad \text{for } 3 \nmid k, \tag{4.52}$$

and, for k odd,

$$12ks(h, k) \equiv k + 1 - 2\left(\frac{h}{k}\right) \pmod{8}, \tag{4.53}$$

with the Legendre-Jacobi symbol on the right-hand side of congruence (4.53) [9].

In our case $k = 5$, $(\lambda, 5) = 1$, we derive from these congruences

$$\frac{1}{2} s(\lambda, 5) + \frac{\lambda - \lambda'}{5} \equiv \frac{1}{4}\left(1 + \left(\frac{\lambda}{5}\right)\right) \pmod{1},$$

and therefore

$$\exp\left(-2\pi i\left(\frac{1}{2}s(\lambda, 5) + \frac{1}{5}(\lambda - \lambda')\right)\right) = -\left(\frac{\lambda}{5}\right).$$

We remark that $(\lambda'/5) = (\lambda/5)$ and obtain from (4.4)

$$(4.6)\qquad 5^{-1/2}\eta(5\tau)^{-1} - \sum_{\lambda=1}^{4}\left(\frac{\lambda}{5}\right)\eta\left(\frac{\tau+24\lambda}{5}\right)^{-1} = 5^{-1/2}\frac{\eta\left(\frac{\tau}{5}\right)^5}{\eta(\tau)^6}.$$

In virtue of (2.2) we can write for the sum on the left-hand side:

$$\begin{aligned}\sum_{\lambda=1}^{4} &= \sum_{\lambda=1}^{4}\left(\frac{\lambda}{5}\right)e^{-\pi i(\tau+24\lambda)/60}\sum_{n=0}^{\infty}p(n)e^{2\pi i n(\tau+24\lambda)/5}\\ &= e^{-\pi i\tau/60}\sum_{n=0}^{\infty}p(n)e^{2\pi i n\tau/5}\sum_{\lambda=1}^{4}\left(\frac{\lambda}{5}\right)e^{(2\pi i\lambda/5)(1-24n)}\\ &= 5^{1/2}e^{-\pi i\tau/60}\sum_{n=0}^{\infty}\left(\frac{n+1}{5}\right)p(n)e^{2\pi i n\tau/5},\end{aligned}$$

where we have evaluated a Gaussian sum, and where for $5\,|\,(n+1)$ the symbol $((n+1)/5)$ means 0, as customary. If we introduce this result into (4.6), apply (2.2) to its first term and finally replace $e^{2\pi i\tau/5}$ by x we obtain

$$(4.7)\qquad \sum_{n=0}^{\infty}p(n)x^{25n} - 5\sum_{n=1}^{\infty}\left(\frac{n}{5}\right)p(n-1)x^n = \frac{\prod(1-x^m)^5}{\prod(1-x^{5m})^6},$$

which is the new identity we wished to derive.

Incidentally, we can construct a formula which is free of infinite products. If we multiply (4.7) by (1.1) the right-hand side will appear as

$$5\frac{1}{\prod(1-x^m)}\cdot\frac{1}{\prod(1-x^{5m})}.$$

These infinite products can be replaced by series by means of (2.1), so that we get

$$(4.8)\qquad \begin{aligned}\left\{\sum_{n=0}^{\infty}p(n)x^{25n} - 5\sum_{n=1}^{\infty}\left(\frac{n}{5}\right)p(n-1)x^n\right\}\sum_{l=0}^{\infty}p(5l+4)x^l\\ = 5\sum_{n=0}^{\infty}p(n)x^n\cdot\sum_{n=0}^{\infty}p(n)x^{5n}.\end{aligned}$$

Comparison of coefficients would yield certain quadratic relations among the $p(n)$.

5. The identity (2.4) can be treated in the same manner. We give only a few highlights. Replacing τ by $-\tau^{-1}$ in (2.4) we have

$$(5.1)\qquad \sum_{\lambda=0}^{6}\eta\left(\frac{24\lambda\tau-1}{7\tau}\right)^{-1} = 7^2\frac{\eta\left(-\frac{7}{\tau}\right)^3}{\eta\left(-\frac{1}{\tau}\right)^4} + 7^3\frac{\eta\left(-\frac{7}{\tau}\right)^7}{\eta\left(-\frac{1}{\tau}\right)^8}.$$

Calling the left-hand member L_7 we rewrite it as

$$(5.2)\qquad L_7 = \eta\left(-\frac{1}{7\tau}\right)^{-1} + \sum_{\lambda=1}^{6} \eta\left(\frac{24\lambda\dfrac{\tau+96\lambda'}{7} + b_\lambda}{7\dfrac{\tau+96\lambda'}{7} - 96\lambda'}\right)^{-1}$$

with

$$(5.21)\qquad \lambda\lambda' \equiv -1 \pmod 7,$$

and

$$(5.22)\qquad b_\lambda = -(24\cdot 96\lambda\lambda' + 1)/7.$$

By means of (4.3) the formula (5.2) goes over into

$$(5.3)\qquad L_7 = (-7i\tau)^{-1/2} + (-i\tau)^{-1/2}\sum_{\lambda=1}^{6} M_\lambda \eta\left(\frac{\tau+96\lambda'}{7}\right)^{-1},$$

where we have

$$M_\lambda = \exp\left\{\pi i\left((s(24\lambda, 7) - \frac{2\lambda - 8\lambda'}{7}\right)\right\}.$$

The congruences (4.51), (4.52), (4.53) yield

$$\frac{1}{2}s(24\lambda, 7) - \frac{\lambda - 4\lambda'}{7} \equiv \frac{1}{4}\left(\frac{\lambda}{7}\right) \pmod 1,$$

and therefore

$$(5.4)\qquad M_\lambda = \exp\left\{\frac{\pi i}{2}\left(\frac{\lambda}{7}\right)\right\} = i\left(\frac{\lambda}{7}\right) = -i\left(\frac{\lambda'}{7}\right).$$

We introduce (5.2), (5.3), (5.4) into (5.1), carry out the modular transformation on the right-hand side, and get thereby

$$(5.5)\qquad \begin{aligned} &7^{-1/2}\eta(7\tau)^{-1} - i\sum_{\lambda'=1}^{6}\left(\frac{\lambda'}{7}\right)\eta\left(\frac{\tau+96\lambda'}{7}\right)^{-1} \\ &\qquad = 7^{1/2}\frac{\eta\left(\dfrac{\tau}{7}\right)^3}{\eta(\tau)^4} + 7^{-1/2}\frac{\eta\left(\dfrac{\tau}{7}\right)^7}{\eta(\tau)^8}. \end{aligned}$$

The sum over λ' can be expanded into an infinite series by means of (2.2):

$$(5.6)\qquad \sum_{\lambda'=1}^{6}\left(\frac{\lambda'}{7}\right)\eta\left(\frac{\tau+96\lambda'}{7}\right)^{-1} = -i7^{1/2}e^{-7\pi i\tau/12}\sum_{n=2}^{\infty}\left(\frac{n}{7}\right)p(n-2)e^{2\pi i n\tau/7};$$

this derivation required the use of the Gaussian sum

$$\sum_{\lambda'=1}^{6}\left(\frac{\lambda'}{7}\right)e^{2\pi\lambda'(96n-4)/7} = -\,i7^{1/2}\left(\frac{n+2}{7}\right).$$

If we now insert (5.6) in (5.5), apply (2.2) and (1.3) in appropriate places, and finally change $e^{2\pi i\tau/7}$ into x we obtain

$$\sum_{n=0}^{\infty} p(n)x^{49n} - 7\sum_{n=2}^{\infty}\left(\frac{n}{7}\right)p(n-2)x^{n} \qquad (5.7)$$
$$= 7x\,\frac{\prod(1-x^{m})^{3}}{\prod(1-x^{7m})^{4}} + \frac{\prod(1-x^{m})^{7}}{\prod(1-x^{7m})^{8}}.$$

6. Up to this moment we have taken the Ramanujan identities (1.1) and (1.2) for granted and have inferred the identities (4.7) and (5.7) as direct consequences. From another point of view, however, we can take these new identities as bases for proofs of (1.1) and (1.2). For that purpose we consider the Ramanujan identities in the forms (2.3) and (2.4), which we prefer to write now as

$$\sum_{\lambda=0}^{4}\eta(5\tau)\eta\left(\frac{\tau+24\lambda}{5}\right)^{-1} = 25\left(\frac{\eta(5\tau)}{\eta(\tau)}\right)^{6}, \qquad (6.1)$$

and

$$\sum_{\lambda=0}^{6}\eta(7\tau)\eta\left(\frac{\tau+24\lambda}{7}\right)^{-1} = 7^{2}\left(\frac{\eta(7\tau)}{\eta(\tau)}\right)^{4} + 7^{3}\left(\frac{\eta(7\tau)}{\eta(\tau)}\right)^{8}. \qquad (6.2)$$

We shall refer to these equations shortly in the abbreviations

$$L_5^*(\tau) = R_5^*(\tau); \qquad L_7^*(\tau) = R_7^*(\tau),$$

respectively.

We can show that $L_5^*(\tau)$ and $R_5^*(\tau)$ are both modular functions of "level 5" ("stufe 5" in Felix Klein's terminology), that is, belonging to a congruence subgroup modulo 5 of the modular group. The subgroup in question is $\Gamma_0(5)$, characterized by $c\equiv 0 \pmod 5$; it is of index 6 in the full modular group. As generators of $\Gamma_0(5)$ we can choose the substitutions (cf. [7, p. 147])

$$S = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \qquad V_2 = \begin{pmatrix}-2 & -1\\ 5 & 2\end{pmatrix}, \qquad V_3 = \begin{pmatrix}-3 & -1\\ 10 & 3\end{pmatrix}$$

and need to test the invariance of $L_5^*(\tau)$ and $R_5^*(\tau)$ only with respect to these 3 substitutions. The discussion of S has already essentially been done in §3. The multiplier $e^{-5\pi i/12}$, which is mentioned there, is exactly absorbed by the factor $\eta(5\tau)$ by which (6.1) differs from (2.3).

As far as $R_5^*(\tau)$ is concerned we have, for V_2,

$$\eta\left(5\,\frac{-\,2\tau\,-\,1}{5\tau\,+\,2}\right) = \eta\left(\frac{-\,2\cdot 5\tau\,-\,5}{5\tau\,+\,2}\right)$$
$$= \exp\left\{-\,\pi i s(-\,2,\,1)\right\}(-\,i(5\tau\,+\,2))^{1/2}\eta(5\tau),$$

in virtue of (4.3), and also

$$\eta\left(\frac{-\,2\tau\,-\,1}{5\tau\,+\,2}\right) = \exp\left\{-\,\pi i s(-\,2,\,5)\right\}(-\,i(5\tau\,+\,2))^{1/2}\eta(\tau).$$

Since $s(-2,1)=0$, as directly seen from (4.31) and (4.32), and

$$s(-\,2,\,5) = -\,s(2,\,5) = 0$$

because of the property of the Dedekind sums

$$s(h,\,k) = 0$$

for

$$h^2 \equiv -\,1 \pmod{k},$$

we have

$$R_5^*(V_2\tau) = R_5^*(\tau).$$

Similarly we find

$$R_5^*(V_3\tau) = R_5^*(\tau).$$

As a matter of fact, not only $R_5^*(\tau)$ but already its sixth root $\eta(5\tau)/\eta(\tau)$ is invariant with respect to V_2 and V_3, but not with respect to S.

The expression $L_5^*(\tau)$ goes over under V_2 into

$$L_5^*(V_2\tau) = \sum_{\lambda=0}^{4}\eta\left(5\,\frac{-\,2\tau\,-\,1}{5\tau\,+\,2}\right)\eta\left(\frac{1}{5}\left(\frac{-\,2\tau\,-\,1}{5\tau\,+\,2}\,+\,24\lambda\right)\right)^{-1}.$$

In order to bring this back into the previous form we need modular substitutions

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

and a summation variable μ such that

$$\frac{1}{5}\left(\frac{-\,2\tau\,-\,1}{5\tau\,+\,2}\,+\,24\lambda\right) = \frac{a\,\dfrac{\tau\,+\,24\mu}{5}\,+\,b}{c\,\dfrac{\tau\,+\,24\mu}{5}\,+\,d}\,.$$

A comparison of the coefficients of the linear functions of τ on both sides leads to

$$\begin{aligned} &\mu = 2 - \lambda, \\ (6.3)\qquad a &= 120\lambda - 2, \qquad b = -24^2\lambda(2-\lambda) - 19, \\ c &= 25, \qquad d = -120(2-\lambda) + 2. \end{aligned}$$

We have therefore:

$$L_5^*(V_2\tau) = \sum_{\lambda=0}^{4} \eta\left(\frac{-2\cdot 5\tau - 5}{5\tau + 2}\right)\eta\left[\frac{a\dfrac{\tau + 24(2-\lambda)}{5} + b}{c\dfrac{\tau + 24(2-\lambda)}{5} + d}\right]^{-1},$$

a, b, c, d being taken from (6.3). Application of (4.3) now shows that

$$L_5^*(V_2\tau) = \sum_{\lambda=0}^{4} M_\lambda \eta(5\tau)\eta\left(\frac{\tau + 24(2-\lambda)}{5}\right)^{-1}$$

with the multiplier

$$M_\lambda = \exp\left\{\pi i\left(s(120\lambda - 2, 25) - \frac{240\lambda - 240}{300}\right)\right\}.$$

Now we have

$$s(120\lambda - 2, 25) = -s(5\lambda + 2, 25),$$

and from (4.51), (4.52), (4.53)

$$12\cdot 25 s(5\lambda + 2, 25) \equiv \begin{cases} 5\lambda + 2 - (-5\lambda + 12) \pmod{25} \\ 0 \pmod{3} \\ 25 + 1 - 2 \pmod{8}, \end{cases}$$

from which we readily derive

$$s(5\lambda + 2, 25) \equiv -\tfrac{4}{5}(\lambda - 1) \pmod{2},$$

so that

$$M_\lambda = 1.$$

Therefore

$$L_5^*(V_2\tau) = L_5^*(\tau)$$

is proved. Reasonings of a similar kind verify the equation

$$L_5^*(V_3\tau) = L_5^*(\tau).$$

7. With $L_5^*(\tau)$ and $R_5^*(\tau)$ the difference

$$D_5(\tau) = L_5^*(\tau) - R_5^*(\tau)$$

belongs also to $\Gamma_0(5)$. If we now can show that $D_5(\tau)$ remains bounded in the whole fundamental region it must be a constant. Now in the interior of the upper τ-half-plane $\eta(\tau)$ is free of poles and zeros and $D_5(\tau)$ is therefore finite. The only parabolic points of the fundamental region of $\Gamma_0(5)$ are the points $\tau = i\infty$ and $\tau = 0$. Now for $\tau \to i\infty$ it is readily seen that $D_5(\tau) \to 0$ since L_5^* and R_5^* tend separately to 0, as their expansions in $e^{2\pi i\tau}$, which can be taken from (1.3), show directly.

In order to test $D_5(\tau)$ for τ near 0 we carry out the substitution $\tau \to -\tau^{-1}$ and study $D_5(-\tau^{-1})$ for τ near $i\infty$. This is now simple with (4.6), which in correspondence to (6.1) we shall have to write as

$$5^{-1/2}\eta\left(\frac{\tau}{5}\right)\eta(5\tau)^{-1} - \sum_{\lambda=1}^{4}\left(\frac{\lambda}{5}\right)\eta\left(\frac{\tau}{5}\right)\eta\left(\frac{\tau+24\lambda}{5}\right)^{-1}$$

$$(7.1) \qquad = 5^{-1/2}\left[\frac{\eta\left(\frac{\tau}{5}\right)}{\eta(\tau)}\right]^6.$$

Indeed, $D_5(-\tau^{-1})$ is the difference of the two members of equation (7.1). In the uniformizing variable $e^{2\pi i\tau/5}$ both members have a pole of the first order at $i\infty$. If therefore the two sides of (7.1) agree in their first term the difference $D_5(-\tau^{-1})$ remains bounded also at $i\infty$ or $D_5(\tau)$ at the second parabolic point $\tau=0$. Instead, however, of comparing the coefficients of the first term of the members of (7.1), it is easier to do it with (4.7). This is equivalent since (7.1) is obtained from (4.7) through the multiplication by

$$5^{-1/2}x^{-1}\prod(1-x^m).$$

Now indeed both sides in (4.7) begin with the term 1.

We have therefore proved that $D_5(\tau)$ is a constant, which can only be zero since $D_5(\tau) \to 0$ with $\tau \to i\infty$ as we have mentioned before. This proves (6.1) and therefore (2.3) and (1.1). Mordell in [2] also proves (6.1) by testing its two members at the parabolic points of $\Gamma_0(5)$. We used here for this purpose the independent theory of $\eta(\tau)$.

8. We can discuss (6.2) in the same manner. First we have to show that $L_7^*(\tau)$ and $R_7^*(\tau)$ are modular functions of level 7, belonging to $\Gamma_0(7)$ with $c \equiv 0 \pmod 7$. This step we could perform in analogy to the procedure in §6 by testing the generating substitutions of $\Gamma_0(7)$ which we can take as

$$S = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \qquad V_3 = \begin{pmatrix}-2 & -1\\ 7 & 3\end{pmatrix}, \qquad V_5 = \begin{pmatrix}-4 & -1\\ 21 & 5\end{pmatrix},$$

(cf. [8]). Such a procedure, however, would not only mean a repetition of previous arguments but would involve a good deal of numerical work, for

which, by the way, the congruences (4.51)–(4.53) would not quite suffice as a basis(²). We prefer therefore to discuss (6.2) on a more general ground, by taking recourse to the following theorems, in which p always designates a prime number greater than 3.

THEOREM 1. *The functions*

$$\Phi_{p,r}(\tau) = \left(\frac{\eta(p\tau)}{\eta(\tau)}\right)^r \tag{8.1}$$

with

$$r(p-1) \equiv 0 \pmod{24} \tag{8.2}$$

have the transformation equation

$$\Phi_{p,r}(V\tau) = \left(\frac{a}{p}\right)^r \Phi_{p,r}(\tau) \tag{8.3}$$

for the modular substitution

$$V\tau = \frac{a\tau + b}{c\tau + d}$$

of $\Gamma_0(p)$, (a/p) *being the Legendre symbol.*

THEOREM 2. *The function*

$$L_p^*(\tau) = \sum_{\lambda=0}^{p-1} \eta(p\tau)\eta\left(\frac{\tau + 24\lambda}{p}\right)^{-1} \tag{8.4}$$

is invariant under the modular substitutions of $\Gamma_0(p)$ *with* $c \equiv 0 \pmod{p}$.

In order not to interrupt the present line of thought we postpone the proof of these theorems to Part III of this paper.

For $p=7$ and $r=4$ and 8 the Theorems 1 and 2 show immediately that $D_7(\tau) = L_7^*(\tau) - R_7^*(\tau)$ as taken from (6.2) is an invariant of $\Gamma_0(7)$. We have now to show that $D_7(\tau)$ remains bounded in the fundamental region of this group. The only parabolic points of that region are again the points $\tau = i\infty$ and $\tau = 0$. For $\tau \to i\infty$ we have $D_7(\tau) \to 0$, since $L_7^*(\tau)$ and $R_7^*(\tau)$ tend separately to 0, in virtue of the factor $e^{\pi i\tau/12}$ before the infinite product in the definition (1.3) for $\eta(\tau)$.

Instead of investigating $D_7(\tau)$ directly for $\tau \to 0$ we carry out the substitution $T\tau = -\tau^{-1}$ and then let τ tend to $i\infty$. But this substitution has been studied in §5. We have therefore $D_7(-\tau^{-1})$ as the difference of the two members of the equation

(²) Cf. Lemmas 1 and 3, §13.

$$
(8.5)\quad
\begin{aligned}
7^{-1/2}\eta\left(\frac{\tau}{7}\right)\eta(7\tau)^{-1} - i\sum_{\lambda'=1}^{6}\left(\frac{\lambda'}{7}\right)\eta\left(\frac{\tau}{7}\right)\eta\left(\frac{\tau+96\lambda'}{7}\right)^{-1}\\
= 7^{1/2}\left(\frac{\eta\left(\frac{\tau}{7}\right)}{\eta(\tau)}\right)^{4} + 7^{-1/2}\left(\frac{\eta\left(\frac{\tau}{7}\right)}{\eta(\tau)}\right)^{8},
\end{aligned}
$$

which is obtained from (5.5) by multiplication with $\eta(\tau/7)$ and which is the result of the transformation of (6.2). The application of (1.3) shows that each member of (8.5) begins with terms in $e^{-4\pi i\tau/7}$, or, in other words has a pole of the second order in the uniformizing variable $e^{2\pi i\tau/7}$. If we can therefore verify that the two members of (8.5) have their pole terms, the first two terms, in common, then the difference $D_7(-\tau^{-1})$ remains bounded also at $\tau = i\infty$, and is bounded in the whole fundamental region. The comparison of the first two terms of each side of (8.5) is much easier to carry out in (5.7), which through multiplication by

$$7^{-1/2}x^{-2}\prod(1-x^m)$$

goes over into (8.5). Now the first two coefficients of both sides of (5.7) are indeed in agreement, they are 1 and 0 for both.

Since therefore $D_7(\tau)$ is bounded in the fundamental region it is a constant, and this constant is obviously 0, since $D_7(\tau)\to 0$ for $\tau\to i\infty$, as mentioned. But $D_7(\tau)=0$ means that the equation (6.2) must hold, and this is equivalent to a proof of (1.2) (cf. [2]).

9. All these reasonings apply also to an identity which Zuckerman [4] has given in the form

$$
(9.1)\quad \sum_{l=0}^{\infty} p(13l+6)x^l = \sum_{j=0}^{6} a_j x^j \frac{\prod(1-x^{13m})^{2j+1}}{\prod(1-x^m)^{2(j+1)}},
$$

where the a_j are certain integers which are computed in Zuckerman's paper. The procedure which we applied to (1.1) and (1.2) in §§4 and 5 leads here to the transformed identity

$$
(9.2)\quad
\begin{aligned}
\sum_{n=0}^{\infty} p(n)x^{169n} - 13\sum_{n=7}^{\infty}\left(\frac{n}{13}\right)p(n-7)x^n\\
= \sum_{j=0}^{6} a_j 13^{1-j}x^{6-j}\frac{\prod(1-x^m)^{2j+1}}{\prod(1-x^{13m})^{2(j+1)}}.
\end{aligned}
$$

The numbers $13^{1-j}a_j$ are integers.

Our method yields now a direct proof of (9.1). We first express (9.1) and (9.2) in terms of $\eta(\tau)$. We have only to observe that $x=e^{2\pi i\tau}$ in (9.1) and $x=e^{2\pi i\tau/13}$ in (9.2). Moreover we multiply the resulting equations by $\eta(13\tau)$ and $\eta(\tau/13)$, respectively.

For the proof of (9.1) we have first to show that

$$\left(\frac{\eta(13\tau)}{\eta(\tau)}\right)^2,$$

and

$$\sum_{\lambda=0}^{12} \eta(13\tau)\eta\left(\frac{\tau+24\lambda}{13}\right)^{-1}$$

belong to the group $\Gamma_0(13)$ with $c\equiv 0 \pmod{13}$. This is at once inferred from the Theorems 1 and 2 for $p=13$ and $r=2$.

Secondly, we simply have to compare the first 7 coefficients of (9.2) since $\eta(\tau/13)\eta(13\tau)^{-1}$ as well as $\eta(\tau/13)^{14}\eta(\tau)^{-14}$ begin with $e^{-14\pi i\tau/13}$, that is, have a pole of 7th order in the uniformizing variable $e^{2\pi i\tau/13}$. Now the comparison of the first seven coefficients of (9.2) yields without too much effort the following seven equations, with $b_j=13^{1-j}a_j$:

$$
\begin{aligned}
1 &= b_6,\\
0 &= -13b_6+b_5,\\
0 &= 65b_6-11b_5+b_4,\\
0 &= -130b_6+44b_5-9b_4+b_3,\\
0 &= -65b_6-55b_5+9b_4-7b_3+b_2,\\
0 &= 728b_6-110b_5-12b_4+14b_3-5b_2+b_1,\\
0 &= -871b_6+484b_5-90b_4+7b_3+5b_2-3b_1+b_0.
\end{aligned}
\tag{9.3}
$$

But these are exactly the equations by which Zuckerman (p. 104 of his paper) determines the coefficients which we here have called a_j. His derivation of these equations is based on an entirely different argument.

Part II. Identities of Watson and Zuckerman

10. G. N. Watson [3] and H. S. Zuckerman [4] have derived identities analogous to those of Ramanujan, but corresponding to powers of 5 and 7 as moduli:

$$\sum_{l=0}^{\infty} p(25l+24)x^l = \sum_{j=1}^{5} b_j x^{j-1}\frac{\prod(1-x^{5m})^{6j}}{\prod(1-x^m)^{6j+1}},\tag{10.1}$$

and

$$\sum_{l=0}^{\infty} p(49l+47)x^l = \sum_{j=1}^{14} c_j x^{j-1}\frac{\prod(1-x^{7m})^{4j}}{\prod(1-x^m)^{4j+1}}.\tag{10.2}$$

The b_j and c_j are integers, which are computed in [4] and which incidentally, in accordance with Ramanujan's theorems about $p(25l+24)$ and $p(49l+47)$, have the properties $25\mid b_j$ and $49\mid c_j$. From our present point of view we can

easily obtain a proof for (10.1) and (10.2). It may be sufficient to carry it out only for the first of these equations.

If in (6.1) we replace τ by $(\tau+24\mu)/5$ we get

$$\sum_{\lambda,\mu=0}^{4} \eta(\tau)\eta\left(\frac{\tau+24\mu+5\cdot 24\lambda}{25}\right)^{-1} = 5^2\sum_{\mu=0}^{4}\left(\frac{\eta(\tau)}{\eta\left(\dfrac{\tau+24\mu}{5}\right)}\right)^6,$$

or

$$\sum_{\lambda=0}^{24} \eta(\tau)\eta\left(\frac{\tau+24\lambda}{25}\right)^{-1} = 5^2\Psi_{5,6}(\tau) \tag{10.31}$$

with

$$\Psi_{5,6}(\tau) = \sum_{\mu=0}^{4}\left(\frac{\eta(\tau)}{\eta\left(\dfrac{\tau+24\mu}{5}\right)}\right)^6. \tag{10.32}$$

Now $\Psi_{5,6}(\tau)$ belongs to the group $\Gamma_0(5)$ as we infer from the following theorem, whose proof we defer to Part III.

THEOREM 3. *The functions*

$$\Psi_{p,r}(\tau) = \sum_{\lambda=0}^{p-1}\left(\frac{\eta(\tau)}{\eta\left(\dfrac{\tau+24\lambda}{p}\right)}\right)^r \tag{10.41}$$

with $r(p-1)\equiv 0 \pmod{24}$ *have the transformation equation*

$$\Psi_{p,r}(V\tau) = \left(\frac{a}{p}\right)^r \Psi_{p,r}(\tau) \tag{10.42}$$

for the substitution

$$V\tau = \frac{a\tau+b}{c\tau+d}$$

with $c\equiv 0 \pmod{p}$, *p being a prime greater than* 3.

We can therefore try to construct $\Psi_{5,6}(\tau)$ as a polynomial in[9]

$$\Phi_{5,6}(\tau) = \left(\frac{\eta(5\tau)}{\eta(\tau)}\right)^6.$$

[9] The background of this possibility is, of course, the fact that (1) $\Phi_{5,6}$ is univalent in the fundamental region, having only a zero of order one at $\tau = i\infty$ and (2) $\Psi_{5,6}$ as well as $\Phi_{5,6}$ are regular in the interior of the fundamental region. However, we do not need this remark, since the following arguments are self-sufficient.

For this purpose we determine coefficients β_j so that

$$\Psi_{5,6}(\tau) = \sum_{j=1}^{N} \beta_j \Phi_{5,6j}(\tau). \tag{10.5}$$

We need to verify this equation only at the two parabolic points of the fundamental region of $\Gamma_0(5)$, namely, at $\tau = i\infty$ and $\tau = 0$. At the former point (10.5) is satisfied since both members tend to 0 as $\tau \to i\infty$. Instead of discussing (10.5) directly for $\tau = 0$ we subject it first to the transformation T, which yields, by the device employed in (4.2),

$$\eta\left(-\frac{1}{\tau}\right)^6 \eta\left(-\frac{1}{5\tau}\right)^{-6} + \sum_{\lambda=1}^{4} \eta\left(-\frac{1}{\tau}\right)^6 \eta\left[\frac{24\lambda \dfrac{\tau + 24\lambda'}{5} + b_\lambda}{5\dfrac{\tau + 24\lambda'}{5} - 24\lambda'}\right]^{-6}$$
$$= \sum_{j=1}^{N} \beta_j \eta\left(-\frac{5}{\tau}\right)^{6j} \eta\left(-\frac{1}{\tau}\right)^{-6j},$$

or, in analogy to (4.6),

$$5^{-3}\left(\frac{\eta(\tau)}{\eta(5\tau)}\right)^6 + \sum_{\lambda=1}^{4}\left(\frac{\eta(\tau)}{\eta\left(\dfrac{\tau + 24\lambda}{5}\right)}\right)^6 = \sum_{j=1}^{N} \beta_j 5^{-3j}\left(\frac{\eta\left(\dfrac{\tau}{5}\right)}{\eta(\tau)}\right)^{6j}. \tag{10.6}$$

Here both members show poles at $\tau = i\infty$: the left-hand member begins with a term in $e^{-2\pi i\tau}$, whereas the right-hand member begins with a term in $e^{-2\pi i\tau N/5}$. We have therefore $N = 5$, that is, a pole of order 5 in the uniformizing variable $e^{2\pi i\tau/5}$.

If now the coefficients β_j, $j = 1, \cdots, 5$, are determined in such a way that both members of (10.6) agree in their pole terms at $\tau = i\infty$, that is, in the 5 first terms, then the difference of the members of (10.5) remains bounded throughout the fundamental region and is therefore a constant, which in particular must be equal to 0. We rewrite (10.6) as

$$\eta(5\tau)^{-6} - 5^3\eta\left(\frac{\tau}{5}\right)^{-6} + 5^3\sum_{\lambda=0}^{4} \eta\left(\frac{\tau + 24\lambda}{5}\right)^{-6} = \sum_{j=1}^{5} \beta_j 5^{3(1-j)} \frac{\eta\left(\dfrac{\tau}{5}\right)^{6j}}{\eta(\tau)^{6(j+1)}},$$

which, if we introduce $e^{2\pi i\tau/5} = x$, and

$$\frac{1}{\prod (1 - x^m)^6} = \sum_{n=0}^{\infty} p_6(n) x^n, \qquad p_6(0) = 1,$$

leads to

$$\sum_{n=0}^{\infty} p_6(n)x^{25n} - 5^3x^6\sum_{n=0}^{\infty} p_6(n)x^n + 5^4x^{10}\sum_{l=0}^{\infty} p_6(5l+4)x^l$$
$$= \sum_{j=1}^{5} \beta_j 5^{-3j+3}x^{5-j}\frac{\prod (1-x^m)^{6j}}{\prod (1-x^{5m})^{6(j+1)}}.$$

We need to ensure only the agreement of the first 5 terms on each side, that is, the terms with $x^0, x, \cdots, x^4$. If we leave aside all unnecessary terms we have therefore to compute $\beta_1, \beta_2, \cdots, \beta_5$ from

$$\sum_{j=1}^{5} \beta_j 5^{-3j+3}x^{5-j}\prod_{m=1}^{4}(1-x^m)^{6j} = 1 + O(x^5).$$

This is equivalent to 5 linear equations for $\beta_j 5^{-3j+3}$, which are solved stepwise, beginning with $\beta_5 5^{-12}=1$. These, however, are exactly the equations which Zuckerman solves on pp. 100, 101 of his paper, and which we do not need to repeat here.

Therefore, the equation (10.5) is proved for $N=5$ and appropriate β_j, $j=1, \cdots, 5$. Comparing (10.5) with (10.31) we obtain

$$\sum_{\lambda=0}^{24} \eta\left(\frac{\tau+24\lambda}{25}\right)^{-1} = 5^2\sum_{j=1}^{5}\beta_j\frac{\eta(5\tau)^{6j}}{\eta(\tau)^{6j+1}}. \tag{10.7}$$

This in turn is equivalent to (10.1) for $b_j=\beta_j$, which therefore is proved.

We remark that for the construction of a similar identity for the modulus 5^3 the required step would be slightly different from that one described at the beginning of this paragraph. We should first have again to replace τ by $(\tau+24\mu)/5$ which would lead to

$$\sum_{\lambda=0}^{124} \eta\left(\frac{\tau+24\lambda}{125}\right)^{-1} = \sum_{j=1}^{5}\beta_j\sum_{\mu=0}^{4}\frac{\eta(\tau)^{6j}}{\eta\left(\dfrac{\tau+24\mu}{5}\right)^{6j+1}}.$$

In order to have modular functions belonging to $\Gamma_0(5)$ we should now have to multiply both sides by $\eta(5\tau)$ (and not by $\eta(\tau)$ as in (10.31)). As a matter of fact, we could prove by theorems analogous to Theorems 1 to 3 that the functions

$$\sum_{\mu=0}^{4}\frac{\eta(\tau)^{6j}\eta(5\tau)}{\eta\left(\dfrac{\tau+24\mu}{5}\right)^{6j+1}},$$

and hence, in virtue of the preceding equation,

$$\sum_{\lambda=0}^{124}\eta(5\tau)\eta\left(\frac{\tau+24\lambda}{125}\right)^{-1}$$

belong to $\Gamma_0(5)$ and therefore admit of a representation as a polynomial in $\Phi_{5,6}(\tau)$. Watson [3], indeed, observes that in an induction from 5^k to 5^{k+1} two different kinds of procedures are required according to the parity of k.

We could discuss and prove (10.2) in a manner completely similar to that applied to (10.1). We refrain from giving the details since no new ideas are involved. It is, moreover, clear that our method proves the existence of similar identities for any power of 13 as modulus.

11. In the preceding paragraph we have applied the substitution T to the right-hand side of (10.7) but not yet to its left-hand side. It is worth while to carry it out since it will lead to a modular equation. With $\tau' = -\tau^{-1}$ we have

$$\sum_{\lambda=0}^{24} \eta\left(\frac{\tau' + 24\lambda}{25}\right)^{-1} = \sum_{\lambda=0}^{24} \eta\left(\frac{24\lambda\tau - 1}{25\tau}\right)^{-1}$$

$$= \sum_{\mu=0}^{4} \eta\left(\frac{24\mu\cdot 5\tau - 1}{5\cdot 5\tau}\right)^{-1} + \sum_{\lambda \bmod 25,\,(\lambda,5)=1} \eta\left(\frac{24\lambda\dfrac{\tau + 24\lambda'}{25} + b_\lambda}{25\dfrac{\tau + 24\lambda'}{25} - 24\lambda'}\right)^{-1} \tag{11.1}$$

where $\lambda\lambda' \equiv -1 \pmod{25}$ and

$$b_\lambda = \frac{1}{25}(-24^2\lambda\lambda' + 1).$$

The first sum on the right side of (11.1) can be taken from (4.1), the second admits the application of (4.3), therefore

$$\sum_{\lambda=0}^{24} \eta\left(\frac{\tau' + 24\lambda}{25}\right)^{-1} = 5^2 \frac{\eta\left(-\dfrac{1}{\tau}\right)^5}{\eta\left(-\dfrac{1}{5\tau}\right)^6}$$

$$+ (-i\tau)^{-1/2} \sum_{\lambda \bmod 25,\,(\lambda,5)=1} M_\lambda \eta\left(\frac{\tau + 24\lambda}{25}\right)^{-1}$$

with

$$M_\lambda = \exp\left\{-\pi i\left(s(24\lambda, 25) - \frac{24\lambda - 24\lambda'}{12\cdot 25}\right)\right\} = 1$$

in virtue of the congruences (4.51), (4.52), (4.53). If we apply the substitution T also to the right-hand side of (10.7) we obtain

$$\frac{1}{5}\frac{\eta(\tau)^5}{\eta(5\tau)} + \sum_{\lambda \bmod 25,\,(\lambda,5)=1} \eta\left(\frac{\tau + 24\lambda}{25}\right)^{-1} = 5^2 \sum_{j=1}^{5} \beta_j 5^{-3j} \frac{\eta\left(\dfrac{\tau}{5}\right)^{6j}}{\eta(\tau)^{6j+1}}. \tag{11.2}$$

We could now, following the procedure of §§4, 5, introduce here infinite series with $p(n)$ as coefficients and infinite products. Instead of doing that we write

$$\sum_{\lambda \bmod 25,\,(\lambda,5)=1} \eta\left(\frac{\tau+24\lambda}{25}\right)^{-1} = \sum_{\lambda=0}^{24} \eta\left(\frac{\tau+24\lambda}{25}\right)^{-1} - \sum_{\mu=0}^{4} \eta\left(\frac{\tau+24\cdot 5\mu}{25}\right)^{-1}.$$

Here we apply (10.7) and (2.2), the latter with $\tau/5$ instead of τ, so that (11.2) goes over into

$$\frac{1}{5}\,\frac{\eta(\tau)^5}{\eta(5\tau)^6} + 5^2\sum_{j=1}^{5}\beta_j\frac{\eta(5\tau)^{6j}}{\eta(\tau)^{6j+1}} - 5^2\frac{\eta(\tau)^5}{\eta\left(\frac{\tau}{5}\right)^6} = 5^2\sum_{j=1}^{5}\beta_j 5^{-3j}\frac{\eta\left(\frac{\tau}{5}\right)^{6j}}{\eta(\tau)^{6j+1}}.$$

If we multiply by $5\eta(\tau)$ and replace τ by 5τ we get

$$\Phi(5\tau)^{-1} - 5^3\Phi(\tau) + 5^3\sum_{j=1}^{5}\beta_j(\Phi(5\tau)^j - 5^{-3j}\Phi(\tau)^{-j}) = 0 \tag{11.3}$$

with

$$\Phi(\tau) = \Phi_{5,6}(\tau) = \left(\frac{\eta(5\tau)}{\eta(\tau)}\right)^6. \tag{11.31}$$

We have already found in §6, and again by means of Theorem 2, that $\Phi(\tau)$ is a modular function of level 5. Therefore (11.3) is a transformation equation of level 5 and order 5. In the form (11.3) it is reducible. We put

$$5^{-3}\Phi(\tau)^{-1} = X, \qquad \Phi(5\tau) = Y, \tag{11.4}$$

and have then, after multiplication of (11.3) by XY,

$$X - Y + 5^3XY\sum_{j=1}^{5}\beta_j(Y^j - X^j) = 0,$$

and after the exclusion of the factor

$$5^3XY\sum_{j=1}^{5}\beta_j\sum_{\nu=0}^{j-1}X^\nu Y^{j-1-\nu} = 1. \tag{11.5}$$

This equation is of degree 5 in Y, as it has to be, because $\Phi(5\tau)$ belongs to the group $\Gamma_0(25)$ with $c\equiv 0 \pmod{25}$, which is of index 5 in the group $\Gamma_0(5)$ of $\Phi(\tau)$. Therefore (11.5) is irreducible.

Moreover the equation (11.5) is symmetric in X and Y. These two functions go over into each other by the substitution

$$\tau' = -(5\tau)^{-1}.$$

With these properties the equation (11.5) fulfills the definition of a modular

equation (Klein-Fricke [5, vol. 2, pp. 56, 57]), it is a modular equation of level 5 and order 5.

There exists also an equation for

$$\frac{\eta(5\tau)}{\eta(\tau)} = (\Phi(\tau))^{1/6} \tag{11.6}$$

which in our notation is

$$5^{5/2}X^{5/6}Y^{1/6}+5^3X^{4/6}Y^{2/6}+3\cdot 5^{5/2}X^{3/6}Y^{3/6}+5^3X^{2/6}Y^{4/6}+5^{5/2}X^{1/6}Y^{5/6}=1 \tag{11.7}$$

(cf. [6, p. 395, formula (23)], and [3, p. 105, formula (3.2)]). The function (11.6) does not belong to $\Gamma_0(5)$, it is in fact a function of level 30. We can look upon (11.7) from our present point of view as on a modular equation of level 5 in "irrational form." By elementary algebraic processes the equation (11.5) can be regained from (11.7).

12. A similar treatment of (10.2) leads to an algebraically different situation. Instead of (11.3) we obtain this time

$$\begin{aligned}&7(\Phi(7\tau)^{-1} - 7^2\Phi(\tau)) + (\Phi(7\tau)^{-2} - 7^4\Phi^2(\tau))\\ &\qquad + 7^3\sum_{j=1}^{14} c_j(\Phi(7\tau)^j - 7^{-2j}\Phi(\tau)^{-j}) = 0\end{aligned} \tag{12.11}$$

with

$$\Phi(\tau) = \Phi_{7,4}(\tau) = \left(\frac{\eta(7\tau)}{\eta(\tau)}\right)^4. \tag{12.12}$$

Theorem 2 shows that $\Phi(\tau)$ is a modular function of level 7, belonging to $\Gamma_0(7)$ and (12.11) is therefore a transformation equation of level 7 and order 7. Let us put

$$7^{-2}\Phi(\tau)^{-1} = X, \qquad \Phi(7\tau) = Y,$$

so that (12.11) can be written as

$$(Y^{-2} - X^{-2}) + 7(Y^{-1} - X^{-1}) + 7^3\sum_{j=1}^{14} c_j(Y^j - X^j) = 0. \tag{12.2}$$

This equation can be freed of the factor $Y-X$. But then it would still remain of degree 15 in Y, whereas $\Phi(7\tau)$ belongs to $\Gamma_0(49)$ which is of index 7 in $\Gamma_0(7)$. Therefore even after division by $Y-X$ equation (12.2) cannot be irreducible. If we write it indeed in the form

$$Y^{-2} + 7Y^{-1} + 7^3\sum_{j=1}^{14} c_jY^j + C = X^{-2} + 7X^{-1} + 7^3\sum_{j=1}^{14} c_jX^j + C \tag{12.3}$$

it turns out by actual computation that for $C=(1/4)7^2\cdot 32145$ and the c_j which are numerically given in Zuckerman's paper [4] we have

$$
\begin{aligned}
&X^{-2} + 7X^{-1} + C + 7^3\sum_{j=1}^{14} c_j X^j \\
&\qquad = (X^{-1} + 7/2 + 82\cdot 7^4 X + 176\cdot 7^6 X^2 + 845\cdot 7^7 X^3 \\
&\qquad\quad + 272\cdot 7^9 X^4 + 46\cdot 7^{11} X^5 + 4\cdot 7^{13} X^6 + 7^{14} X^7)^2,
\end{aligned}
\tag{12.4}
$$

a complete square. This permits the extraction of the square root in (12.3). The square root must be taken with the same sign on both sides so that the constant term disappears and a further division by $X-Y$ can reach the degree 7 of the equation which before that division is

$$
\begin{aligned}
&(Y^{-1} - X^{-1}) + 82\cdot 7^4(Y - X) + 176\cdot 7^6(Y^2 - X^2) + 845\cdot 7^7(Y^3 - X^3) \\
&\quad + 272\cdot 7^9(Y^4 - X^4) + 46\cdot 7^{11}(Y^5 - X^5) + 4\cdot 7^{13}(Y^6 - X^6) \\
&\quad + 7^{14}(Y^7 - X^7) = 0.
\end{aligned}
\tag{12.5}
$$

Multiplication by XY and then division by $Y-X$ establish an equation which is symmetric in X and Y and of degree 7 in Y and therefore irreducible, since 7 is the index of the group Y in that of X, as mentioned above. This equation is therefore a modular equation of level 7 and order 7. Again, an "irrational form" of it is known for

$$
\frac{\eta(7\tau)}{\eta(\tau)} = (\Phi(\tau))^{1/4}. \tag{12.6}
$$

It was given by Watson [3, p. 118, (5.2)] and appears in our notation as

$$
\begin{aligned}
&7^{7/2}X^{7/4}Y^{1/4} + 7^4X^{6/4}Y^{2/4} + 3\cdot 7^{7/2}X^{5/4}Y^{3/4} \\
&\quad + 7^4X^{4/4}Y^{4/4} + 3\cdot 7^{7/2}X^{3/4}Y^{5/4} + 7^4X^{2/4}Y^{6/4} + 7^{7/2}X^{1/4}Y^{7/4} \\
&\quad + 7^{5/2}X^{3/4}Y^{1/4} + 5\cdot 7^2X^{2/4}Y^{2/4} + 7^{5/2}X^{1/4}Y^{3/4} = 1.
\end{aligned}
\tag{12.7}
$$

The equation (12.5) has as roots the 4th powers of the roots of (12.7) and can therefore (disregarding the factor $Y-X$ in (12.5)) be directly derived from (12.6). A control by numerical calculation shows indeed complete agreement in the coefficients. The function (12.6), by the way, is of level 28.

Part III. Proofs for the Theorems 1, 2, and 3

13. The proofs will be based on three lemmas, which occur as Theorems 17, 18, 19 in [9].

Lemma 1. *Let $\Theta = \Theta_k$ denote 1 for $3 \nmid k$ and 3 for $3 \mid k$ so that $\Theta\cdot k$ is either prime to 3 or divisible by 3^2. For $(h, k) = 1$ we have*

$$
12hks(h, k) \equiv h^2 + 1 \pmod{\Theta\cdot k}. \tag{13.1}
$$

Moreover

$$
12ks(h, k) \equiv 0 \pmod 3, \qquad \text{if } 3 \nmid k. \tag{13.2}
$$

LEMMA 2. *For odd k we have*

$$12ks(h, k) \equiv k + 1 - 2\left(\frac{h}{k}\right) \pmod 8, \tag{13.3}$$

where (h/k) denotes the Legendre-Jacobi symbol.

LEMMA 3. *If k is equal to $2^\lambda l$, $\lambda \geqq 0$ and l and h are odd integers, then*

$$12hks(h, k) \equiv h^2 + k^2 + 3k + 1 + 2k\left(\frac{k}{h}\right) \pmod{2^{\lambda+3}}. \tag{13.4}$$

We derive first two further lemmas about the "Dedekind sums" $s(h, k)$. In the sequel p will always be a prime number greater than 3, and r is an integer such that

$$r(p - 1) \equiv 0 \pmod{24}. \tag{13.51}$$

The condition imposed on p will be used in the form

$$p^2 \equiv 1 \pmod{24}. \tag{13.52}$$

LEMMA 4. *Let a, b, c, d be integers with $ad-bc=1$, $c>0$, $p|c$. Put $c=p\cdot c_1$ and*

$$G = \left(s(a, c) - \frac{a+d}{12c}\right) - \left(s(a, c_1) - \frac{a+d}{12c_1}\right). \tag{13.61}$$

Then with r satisfying (13.51) *we have*

$$rG \equiv \frac{1}{2}\left\{1 - \left(\frac{a}{p}\right)^r\right\} \pmod 2. \tag{13.62}$$

Proof. From (13.1) we obtain

$$12ac\left(s(a, c) - \frac{a+d}{12c}\right) \equiv a^2 + 1 - a(a + d) \equiv - bc \pmod{\Theta c},$$

where $\Theta = \Theta_c$ is defined as in Lemma 1. If we then apply (13.1) with $k=c/p=c_1$, we have, after multiplication by p,

$$12ac\left(s(a, c_1) - \frac{a+d}{12c_1}\right) \equiv pa^2 + p - pa(a + d) \equiv - pbc \pmod{\Theta c}.$$

It has to be remarked that Θ has the same value in both congruences, since c and c_1 have either both the factor 3 or not. With the definition (13.61) we obtain

$$12acr\, G \equiv r(p-1)\, bc \equiv 0 \pmod{\Theta c}.$$

For $3\nmid c$ we have $12crG\equiv 0 \pmod 3$ as direct consequence from (13.2) and (13.51). Therefore

$$12crG \equiv 0 \pmod{3c}. \tag{13.71}$$

Now let c be odd. Then Lemma 2 is applicable and yields for $k=c$

$$12c\left(s(a,c) - \frac{a+d}{12c}\right) \equiv c + 1 - 2\left(\frac{a}{c}\right) - (a+d) \pmod 8,$$

and for $k=c/p=c_1$

$$12c\left(s(a,c_1) - \frac{a+d}{12c_1}\right) \equiv c + p - 2p\left(\frac{a}{c_1}\right) - p(a+d) \pmod 8.$$

We can replace (a/c_1) by $(a/c_1)(a/p^2)=(a/cp)=(a/c)(a/p)$, and get therefore

$$\begin{aligned} 12crG &\equiv r(1-p) - 2r\left(\frac{a}{c}\right)\left\{1 - p\left(\frac{a}{p}\right)\right\} + r(p-1)(a+d) \\ &\equiv \pm 2r\left\{1 - p\left(\frac{a}{p}\right)\right\} \pmod 8, \end{aligned}$$

which implies

$$12crG \equiv 0 \pmod 8 \tag{13.72}$$

for r even. For r odd the condition (13.51) necessitates $p\equiv 1 \pmod 8$ and therefore

$$1 - p\left(\frac{a}{p}\right) \equiv 1 - \left(\frac{a}{p}\right) \pmod 4,$$

so that $(a/p)=1$ leads also to (13.72). For r odd together with $(a/p)=-1$ we find, however,

$$12crG \equiv 4 \pmod 8,$$

or

$$12crG \equiv 12c \pmod 8. \tag{13.73}$$

The congruences (13.71), (13.72), (13.73) prove (13.62) for odd c.

For even $c=2^\lambda\gamma$ with γ odd we make use of Lemma 3 and obtain for $k=c$ and $k=c/p=c_1$

$$\begin{aligned} 12ac\left(s(a,c) - \frac{a+d}{12c}\right) &\equiv a^2 + c^2 + 3c + 1 + 2c\left(\frac{c}{a}\right) - a(a+d) \\ &\equiv c^2 + 3c - bc + 2c\left(\frac{c}{a}\right) \pmod{2^{\lambda+3}}, \end{aligned}$$

and

$$12ac\left(s(a, c_1) - \frac{a+d}{12c_1}\right) \equiv pa^2 + \frac{c^2}{p} + 3c + p + 2c\left(\frac{c_1}{a}\right) - pa(a+d)$$
$$\equiv \frac{c^2}{p} + 3c - pbc + 2c\left(\frac{c}{a}\right)\left(\frac{p}{a}\right) \pmod{2^{\lambda+3}},$$

and hence

$$12acrG \equiv r\frac{c^2}{p}(p-1) + r(p-1)bc + 2rc\left(\frac{c}{a}\right)\left\{1 - \left(\frac{p}{a}\right)\right\}$$
$$\equiv \pm 2^{\lambda+1}r\left\{1 - \left(\frac{p}{a}\right)\right\} \pmod{2^{\lambda+3}},$$

and therefore

$$12crG \equiv 0 \pmod{2^{\lambda+3}} \tag{13.74}$$

for r even, as well as for $(p/a)=1$, which for the case of odd r and thus $p\equiv 1 \pmod 4$ is equivalent to $(a/p)=1$.

For r odd in conjunction with $(a/p)=-1$ we have

$$12crG \equiv 2^{\lambda+2} \pmod{2^{\lambda+3}}. \tag{13.75}$$

The congruences (13.71), (13.74), (13.75) establish (13.62) also for the case of an even $c=2^\lambda\gamma$. This finishes the proof of Lemma 4.

LEMMA 5. *Let a, b, c, d be integers with $ad-bc=1$, $c>0$, $p^2|c$. Then, with $c=p^2c_2$, we have*

$$H = \left(s(a, c) - \frac{a+d}{12c}\right) - \left(s(a, c_2) - \frac{a+d}{12c_2}\right) \equiv 0 \pmod 2. \tag{13.8}$$

Proof. For $k=c$ Lemma 1 yields

$$12ac\left(s(a, c) - \frac{a+d}{12c}\right) \equiv a^2 + 1 - a(a+d) \equiv -bc \pmod{\Theta c},$$

and for $k=c/p^2=c_2$

$$12ac\left(s(a, c_2) - \frac{a+d}{c_2}\right) \equiv p^2(a^2+1) - p^2a(a+d) \equiv -p^2bc \pmod{\Theta c}.$$

Therefore

$$12cH \equiv 0 \pmod{\Theta c}.$$

This congruence holds, modulo $3c$, also if $3\nmid c$ since then according to (13.2)

$12c\,s(a,c)$ and $12c\,s(a,c_2)$ are separately divisible by 3, and so is $(p^2-1)a(a+d)$. Thus we obtain

$$12cH \equiv 0 \pmod{3c}. \tag{13.91}$$

Now suppose, first, c to be odd. Then we infer from Lemma 2

$$12c\left(s(a,c) - \frac{a+d}{12c}\right) \equiv c + 1 + 2\left(\frac{a}{c}\right) - (a+d) \pmod 8,$$

and

$$12c\left(s(a,c_2) - \frac{a+d}{12c_2}\right) \equiv c + p^2 + 2p^2\left(\frac{a}{c_2}\right) - p^2(a+d) \pmod 8.$$

But since $(a/c_2) = (a/c)$ and $p^2 - 1 \equiv 0 \pmod 8$ we find by subtraction

$$12cH \equiv 0 \pmod 8, \tag{13.92}$$

which together with (13.91) proves (13.8) for odd c.

Secondly, in case we have $c = 2^\lambda\gamma$, γ and a odd, we obtain from Lemma 3

$$\begin{aligned} 12ac\left(s(a,c) - \frac{a+d}{12c}\right) &\equiv a^2 + c^2 + 3c + 1 + 2c\left(\frac{c}{a}\right) - a(a+d) \\ &\equiv c^2 + 3c - bc + 2c\left(\frac{c}{a}\right) \pmod{2^{\lambda+3}}, \end{aligned}$$

and

$$\begin{aligned} 12ac\left(s(a,c_2) - \frac{a+d}{12c_2}\right) &\equiv p^2a^2 + \frac{c^2}{p^2} + 3c + p^2 + 2c\left(\frac{c_2}{a}\right) - p^2a(a+d) \\ &\equiv \frac{c^2}{p^2} + 3c - p^2bc + 2c\left(\frac{c}{a}\right) \pmod{2^{\lambda+3}}, \end{aligned}$$

and therefore

$$12acH \equiv \frac{c^2}{p^2}(p^2-1) + (p^2-1)bc \equiv 0 \pmod{2^{\lambda+3}},$$

or

$$12cH \equiv 0 \pmod{2^{\lambda+3}}. \tag{13.93}$$

This together with (13.91) completes the proof of (13.8) for c even.

14. **Proof of Theorem** 1. Here and in the following proof we treat the modular substitutions with $c=0$ separately. Since all of these are iterations of $S\tau = \tau + 1$, it suffices in this case to study only S. Now the definitions (8.1) and (1.3) show immediately

$$\Phi_{p,r}(\tau+1) = e^{\pi i r(p-1)/12}\Phi_{p,r}(\tau)$$

which because of (8.2) amounts to

$$\Phi_{p,r}(\tau + 1) = \Phi_{p,r}(\tau).$$

This can be subsumed under (8.3) for $a=1$. From now on we can assume that $c>0$, with $c\equiv 0 \pmod{p}$. The formula (4.3) yields here

$$\Phi_{p,r}(V\tau) = M_V \cdot \Phi_{p,r}(\tau)$$

with

$$M_V = \exp\left\{\pi i r\left(s(a, c) - \frac{a+d}{12c} - s(a, c_1) + \frac{a+d}{12c_1}\right)\right\}, \qquad c_1 = \frac{c}{p}.$$

Application of Lemma 4 gives

$$M_V = \exp\left\{\frac{\pi i}{2}\left(1 - \left(\frac{a}{p}\right)\right)\right\}.$$

This can be written briefly as

$$M_V = \left(\frac{a}{p}\right)^r,$$

which proves Theorem 1.

15. **Proof of Theorem** 2. We first consider the effect of S on $L_p^*(\tau)$. The definitions (8.4) and (1.3) show that in the sum defining $L_p^*(\tau)$ it is only essential that λ runs through a complete residue system modulo p. We can therefore write

$$L_p^*(\tau) = \sum_{\lambda=0}^{p-1} \eta(p\tau)\eta\left(\frac{\tau + 24N + 24\lambda}{p}\right)^{-1}$$

for any integer N. Hence we get

$$L_p^*(\tau + 1) = \sum_{\lambda=0}^{p-1} \eta(p(\tau + 1))\eta\left(\frac{\tau + 24N + 1 + 24\lambda}{p}\right)^{-1}.$$

If we choose here

$$24N + 1 = p^2,$$

we obtain

$$L_p^*(\tau + 1) = \sum_{\lambda=0}^{p-1} \eta(p\tau + p)\eta\left(\frac{\tau + 24\lambda}{p} + p\right)^{-1},$$

which is equal to $L_p^*(\tau)$ in virtue of (1.3). Having thus disposed of the case $c=0$ we can suppose from now on $c>0$ with $c\equiv 0 \pmod{p}$.

In

$$V = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

we have therefore necessarily $a \not\equiv 0 \pmod{p}$. Since only a complete residue system of the index of summation was essential we can also write

$$L_p^*(\tau) = \sum_{\lambda=0}^{p-1} \eta(p\tau)\cdot\eta\left(\frac{\tau + (p^2-1)a\lambda}{p}\right)^{-1}.$$

We obtain therefore, with $c = pc_1$,

$$L_p^*(V\tau) = \sum_{\lambda=0}^{p-1} \eta\left(\frac{ap\tau + pb}{c_1 p\tau + d}\right)\cdot\eta\left(\frac{1}{p}\left(\frac{a\tau+b}{c\tau+d} + (p^2-1)a\lambda\right)\right)^{-1}. \tag{15.1}$$

Proceeding here as we did in §6 for $p=5$ we wish to construct modular substitutions

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

such that

$$\frac{1}{p}\left(\frac{a\tau+b}{c\tau+d} + (p^2-1)a\lambda\right) = \frac{A(\tau + (p^2-1)d\mu)/p + B}{C(\tau + (p^2-1)d\mu)/p + D}, \tag{15.2}$$

where A, B, C, D and μ will depend on λ.

Comparison of coefficients shows that the equations

$$\begin{aligned} a + c(p^2-1)a\lambda &= A, & b + d(p^2-1)a\lambda &= pB + A(p^2-1)d\mu, \\ pc &= C, & pd &= pD + C(p^2-1)d\mu \end{aligned} \tag{15.3}$$

are necessary and sufficient. It is clear that for any choice of λ and μ the numbers A, B, C, D are uniquely determined through (15.3) and satisfy

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = 1. \tag{15.4}$$

We can now tie μ to λ in such a way that B will become an integer, whereas A, C, D are obviously integers for any integers λ, μ. Since p divides c we have

$$a \equiv A \pmod{p},$$

and therefore

$$b - ad\lambda \equiv -ad\mu \pmod{p},$$

which implies already that B is an integer. Now

$$1 = ad - bc \equiv ad \pmod{p},$$

and therefore

$$\mu \equiv \lambda - b \pmod{p}.$$

The new summation index μ needs only to be determined modulo p and we can therefore, without loss of generality, put

$$\mu = \lambda - b. \tag{15.5}$$

This choice now completes the determination of A, B, C, D, which become

$$\begin{aligned} A &= a + c(p^2 - 1)a\lambda, & B &= padb - \frac{c}{p}b^2 - \frac{c}{p}(p^2 - 1)^2 ad\lambda(\lambda - b), \\ C &= pc, & D &= d - c(p^2 - 1)d(\lambda - b). \end{aligned} \tag{15.6}$$

From (15.1), (15.2), and (4.3) we derive now

$$L_p^*(V\tau) = \sum_{\lambda=0}^{p-1} M_\lambda \cdot \eta(p\tau) \cdot \eta\left(\frac{\tau + (p^2 - 1)d\mu}{p}\right)^{-1} \tag{15.7}$$

with

$$M_\lambda = \exp\left\{-\pi i\left(\left(s(a, c_1) - \frac{a + d}{12c_1}\right) - \left(s(A, C) - \frac{A + D}{12C}\right)\right)\right\}.$$

From (15.6) it follows that

$$a \equiv A \pmod{c}.$$

This permits us to write

$$s(a, c_1) = s(A, c_1) = s(A, C_2), \qquad C_2 = \frac{C}{p^2} = \frac{c}{p} = c_1.$$

Moreover the equations (15.6) show that $(a+d)/12c$, and $(A+D)/12c$ differ only by an even integer. We can therefore write

$$M_\lambda = \exp\left\{-\pi i\left(\left(s(A, C_2) - \frac{A + D}{12C_2}\right) - \left(s(A, C) - \frac{A + D}{12C}\right)\right)\right\},$$

and obtain then from Lemma 4

$$M_\lambda = 1.$$

If we observe that in (15.7) μ runs with λ through a complete residue system modulo p we have proved

$$L_p^*(V\tau) = L_p^*(\tau),$$

which finishes the proof of Theorem 2.

16. **Proof of Theorem 3.** The proof is closely similar to that of Theorem 2. The same auxiliary substitutions are used. It is then only necessary to apply Lemma 4 instead of Lemma 5 for the computation of the multiplier M_λ.

REFERENCES

1. S. Ramanujan, *Some properties of $p(n)$, the number of partitions of n*, Collected Papers, pp. 210–213.

2. L. J. Mordell, *Note on certain modular relations considered by Messrs. Ramanujan, Darling, and Rogers*, Proceedings of the London Mathematical Society, (2), vol. 20 (1922), pp. 408–416.

3. G. N. Watson, *Ramanujans Vermutung über Zerfällungsanzahlen*, Journal für die reine und angewandte Mathematik, vol. 179 (1938), pp. 97–128.

4. H. S. Zuckerman, *Identities analogous to Ramanujan's identities involving the partition function*, Duke Mathematical Journal, vol. 5 (1939), pp. 88–110.

5. F. Klein und R. Fricke, *Vorlesungen über die Theorie der Elliptischen Modulfunktionen*, vol. 1, 1890, and vol. 2, 1892.

6. R. Fricke, *Die Elliptischen Funktionen und ihre Anwendungen*, vol. 1, 1916, and vol. 2, 1922.

7. H. Rademacher, *Über die Erzeugenden von Kongruenzuntergruppen der Modulgruppe*, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 7 (1930), pp. 134–148.

8. ———, *Zur Theorie der Modulfunktionen*, Journal für die reine und angewandte Mathematik, vol. 167 (1931), pp. 312–336.

9. H. Rademacher and A. Whiteman, *On Dedekind sums*, American Journal of Mathematics, vol. 63 (1941), pp. 377–407.

UNIVERSITY OF PENNSYLVANIA,
PHILADELPHIA, PA.

THE STRUCTURE OF LOCALLY CONNECTED TOPOLOGICAL SPACES

BY

G. E. ALBERT AND J. W. T. YOUNGS

Introduction

0.1. This paper presents an investigation of the following problem. Exhibit a class $\mathfrak{X}$ of topological spaces which contains all peano spaces and which has the following properties: (1) a cyclic element theory exists in each space of the class, (2) the abstract set consisting of all cyclic element of any space X of the class can be topologized so as to be a member of the class $\mathfrak{X}$, and (3) the hyperspace thus obtained is acyclic. Since the class $\mathfrak{P}$ of all peano spaces does not satisfy the condition (2), it is clear that any solution of the problem lies in a generalization of peano spaces. (For cyclic element theory, see Whyburn [5](1), and [6], or Kuratowski and Whyburn [4].)

One such generalization has been proposed by R. L. Moore [5]; another by one of the authors (Youngs [9]). Moore employed two primitive concepts: region and contiguity (compare this with satelliticity 4.8). Youngs used the notion of arc as primitive. In the following pages a solution is given which is based upon the usual concept of open set.

0.2. The work is divided into four sections. The first of these is devoted to the definition and a brief discussion of the class $\mathfrak{X}$ of spaces to be used in the remainder of the paper; namely, locally connected topological spaces. No separation or countability properties are assumed. Thus, in particular, a single point need not form a closed point set. The use of such a weak topology is not dictated merely by a desire for generality; indeed, it is shown in later sections that this weakness is fundamental in the consideration of hyperspaces of peano spaces.

The development of a theory of cyclic elements for spaces of the class $\mathfrak{X}$ occupies the second section. In such spaces, the standard definitions of "cut point" and "cyclic element" (Kuratowski and Whyburn [4]) fail to yield certain important properties of these concepts. However, the properties are easily recovered by generalizations of the definitions mentioned. The degree of similarity achieved between the cyclic structure of spaces of the general class $\mathfrak{X}$ and of peano spaces seems remarkable in view of the weak topology assumed in the former.

Presented to the Society in part on April 12, 1940 by the first of the authors under the title *On contiguous point spaces and their applications*, and in final form under the present title on April 12, 1941; received by the editors April 30, 1941.

(1) Numbers in brackets refer to the bibliography.

The third section of the paper contains a discussion of the hyperspace of cyclic elements of any locally connected topological space. In particular, it is shown that the class of all such spaces solves the problem stated in 0.1. Moreover, it is shown that the hyperspace can always be defined as the strongly continuous image (see 3.4) of the original space.

0.3. It is easily seen that the solution offered for the problem of 0.1 is not unique. The concluding section of the paper is devoted to a discussion of certain subclasses of $\mathfrak{X}$ which also solve the problem. In this study a new concept, that of hereditary classes of spaces, arises naturally. Briefly, a subclass $\mathcal{H}$ of $\mathfrak{X}$ is called hereditary if, whenever X is in $\mathcal{H}$, the hyperspace of X is in $\mathcal{H}$ and every true cyclic element of X is a member of $\mathcal{H}$. The class $\mathcal{P}$ of all peano spaces is not hereditary. However, $\mathfrak{X}$ is a hereditary class that contains $\mathcal{P}$. It is interesting to note that there are highly restricted hereditary subclasses of $\mathfrak{X}$ which contain the class $\mathcal{P}$. For example, one such subclass is composed of all the locally connected topological spaces which satisfy the T_0 separation axiom, and which are, in addition, strongly continuous images of the closed unit interval of the number axis. This class of spaces yields considerable insight into the nature of the hyperspaces of peano spaces.

I. The space X

1.1. The symbol X will denote a class of elements (the space); elements of X, to be called points, will be denoted by small Latin letters; point sets will be designated by capital Latin letters. The usual logical concepts and notations will be employed in dealing with point sets.

It will be supposed that there is defined in X a definite class $\mathfrak{O}$ of point sets. A set will be called *open* if and only if it belongs to $\mathfrak{O}$. The collection X is called a *topological space* if:

A1. The class X and the empty set are open.

A2. The product of any two open sets is open.

A3. The sum of any number of open sets is open.

Only spaces of this character will be considered.

1.2. A point set will be called *closed* if it is the complement of an open set.

1.3. If A is a point set in X, the symbol $\overline{A}$ will denote the product of all the closed subsets of X that contain A. The set $\overline{A}$ will be called the *closure* of the set A.

1.4. A space in which the closure $\overline{A}$ of an arbitrary set A is the primitive concept is called a Kuratowski space whenever the closure function satisfies the axioms:

$$\overline{A+B}=\overline{A}+\overline{B}, \qquad A\subset\overline{A}, \qquad \overline{\overline{A}}\subsetneq\overline{A}, \qquad \overline{0}=0.$$

A topological space is completely equivalent to a Kuratowski space. The proof of this theorem is well known (Alexandroff and Hopf [1, p. 41] or Hausdorff

[2, p. 227]) and will not be reproduced here. The following consequence of the result will be used in the sequel without further mention.

If p is a point and A is a set, then $p \in \bar{A}$ if and only if every open set that contains p intersects A.

1.5. The symbol $F(A)$ will denote the set

$$\bar{A} \cdot \overline{(X - A)}$$

which is called the *frontier* of the set A. It is easily seen that the frontier of any set is closed.

1.6. A point set E will be called *connected* if it cannot be decomposed into two nonvacuous sets A and B such that $A \cdot \bar{B} = 0 = \bar{A} \cdot B$. Any degenerate set (the empty set or any set consisting of a single point) is connected. If E is connected and $E \subset C \subset \bar{E}$, then C is connected. (For other results on connected sets in general spaces see Knaster and Kuratowski [3].)

1.7. If $S \neq 0$ is a maximal connected subset of E, then S is called a *component* of E.

1.8. The topological space X will be called *locally connected* if each component of any open set is open. This definition is equivalent to the usual one in peano spaces.

In the remainder of the paper it will be assumed that X is a locally connected topological space.

1.9. LEMMA. *If G_1 and G_2 are open, disjoint point sets, then $G_1 \cdot \bar{G}_2 = 0 = \bar{G}_1 \cdot G_2$, and $\bar{G}_1 \cdot \bar{G}_2 \subset X - (G_1 + G_2)$.*

1.10. THEOREM. *If E is a closed point set and S is any component of the set $X - E$, then $F(S) \subset E$.*

Proof. Let R be the sum of all the components of $X - E$ which are distinct from S. Since S is open $F(S) \subset \overline{X - S} = X - S$. On the other hand, $F(S) \subset \bar{S}$ and $\bar{S}R = 0$ since S is a component of $S + R$. Therefore, $F(S) \subset X - S - R = E$.

II. Cyclic elements

2.1. In this section a definition will be given for a class of subsets of the space X which will be called cyclic elements. The concept as introduced here is easily shown to reduce to the familiar one in case X is a peano space.

The structure of a locally connected topological space will depend only upon the structure of its individual components. Thus, there will be no loss in generality in assuming that the space X is connected and this will be done. The restriction will be removed in 3.18.

2.2. A point p will be said to *separate* two points a and b if: (1) $\bar{p} = p$, and (2) the points a and b lie in distinct components of the set $X - p$. Any point of this character will be called a *cut point* of the space X.

The exclusion of all points which are not closed from the class of cut points

is not an artificial condition. The reader will recall that in peano spaces the essential tool in the use of cut points is the property that the components of the complement of any cut point are open sets. This property would be lost here without the condition (1) on cut points.

2.3. Two points a and b will be called *conjugate* (denoted by $a\sim b$) if they are separated by no point of X. Conjugacy is reflexive and symmetric.

Note that, as the term separate does not apply to a point p such that $\bar{p}\neq p$, two points a and b may be conjugate even though they lie in distinct components of the set $X-p$.

2.4. Lemma. *If $a\sim x_1\sim \cdots \sim x_n\sim b$, and the point z separates a from b, then $z=x_i$ for some $i=1, \cdots, n$*(2).

Proof. If the statement is false, then all the points $a, x_1, \cdots, x_n, b$ lie in a single component of the set $X-z$.

2.5. Corollary. *If $a\sim x\sim b$ and $\bar{x}\neq x$, then $a\sim b$.*

2.6. Corollary. *If $a\sim x_1\sim \cdots \sim x_n\sim b$, $a\sim y_1\sim \cdots \sim y_m\sim b$ and all the points $x_1, \cdots, x_n, y_1, \cdots, y_m$ are distinct, then $a\sim b$.*

2.7. A point set E will be called *coherent* if for every pair of points a and b contained in E it is true that $a\sim b$(2).

2.8. A point set E will be called *complete* if E contains every point z which is conjugate to each of two distinct points contained in E(2).

2.9. Lemma. *If E is any coherent set and x is any point such that $\bar{x}=x$, then the set $E-x$ is contained in a single component of the set $X-x$.*

2.10. A point set N which is nondegenerate, complete, and coherent will be called an *N-set.*

2.11. Theorem. *If $a\sim b$ and $a\neq b$, then there exists a unique N-set containing a and b.*

Proof. Denote by N the totality of points which are conjugate to both a and b. If x and y are any pair of distinct points in the set N, then by 2.6, $x\sim y$. Thus N is a coherent set. Suppose that z is a point conjugate to both x and y. Since $x\sim a\sim y$, it follows by 2.6 that $z\sim a$. Similarly, $z\sim b$. Thus z is in the set N, and N is a complete set. By definition N is an N-set.

If N' is any other N-set containing $a+b$, it follows directly from the completeness of the sets N and N' that they are identical.

2.12. Theorem. *If N_1 and N_2 are distinct N-sets, then $N_1\cdot N_2$ is either vacuous or a single point x which separates any point of N_1-x from every point of N_2-x.*

---

(2) The importance of Lemma 2.4 and the definitions in 2.7 and 2.8 were noticed first by Radó and Reichelderfer [8].

Proof. By 2.11, the product $N_1 \cdot N_2$ is either vacuous or a point x. Let $a_i \neq x$ be in N_i for $i=1, 2$. If $\bar{x} \neq x$, then by 2.5, $a_1 \sim a_2$. If $\bar{x}=x$ but x does not separate a_1 from a_2, then by 2.4, $a_1 \sim a_2$. In either case $a_1 \sim a_2 \sim x$ and thus by the completeness of N_1 it contains a_2. This contradicts 2.11.

2.13. THEOREM. *Any N-set is closed.*

Proof. Suppose that the point $p \in \overline{N}$. Choose any point $a \in N$. If there is a point x such that $\bar{x}=x$ and $a \neq x \neq p$, let S be the component of $X-x$ containing p. Now S is open by 1.8, and so contains a point of the set N. As N is coherent, S contains the point a. Thus $p \sim a$. If no such point x exists, $p \sim a$ by the definition of 2.2. But the point a was arbitrarily chosen in N. Since N is nondegenerate and complete, $p \in N$.

Remark. In connection with the remarks of 2.2, the theorem of the present section would be false without the generalized definition of cut points introduced in 2.2.

2.14. Consider any N-set, N. Define the subset $k(N)$ of N by: the point x is in $k(N)$ if and only if (1) $x \in N$, and (2) for no component S of the set $X-N$ is it true that $x \in F(S)$. The set $k(N)$ will be called the ***kernel*** of the set N.

Clearly, by 1.10, $N=k(N)+\sum F(S)$ where the summation extends over all components S of the set $X-N$.

2.15. A subset M of the space X will be termed a ***true cyclic element*** if: (1) it is an N-set, (2) the set S being any component of $X-M$, the frontier $F(S)$ is a single point, and (3) the kernel $k(M)$is nondegenerate. The symbol M will be employed as a generic notation for true cyclic element.

2.16. If the space X is peanian, it can be shown that every N-set given by the definition of 2.10 is a true cyclic element under the definition of 2.15. Moreover, since every point in a peano space is a closed set, the definitions of cut point and conjugate points given in 2.2 and 2.3 reduce to the standard notions for such spaces. It follows readily that, for peano spaces, the definition of true cyclic elements given here reduces to the usual definition (Kuratowski and Whyburn [4]).

2.17. A point which is not a cut point of the space and which is contained in no true cyclic element will be called ***singular***. Cut points and singular points will be termed degenerate cyclic elements. It follows immediately that:

2.18. THEOREM. *The cyclic elements of the space X cover it.*

2.19. THEOREM. *If M is a true cyclic element and x is a point of M, then $x \in M-k(M)$ if and only if x is a cut point of the space X.*

Proof. Suppose x is a cut point of X. Using 2.9 select a component S of $X-x$ such that $S \cdot M=0$. Then S is also a component of $X-M$. But $x=F(S)$, so $x \in M-k(M)$.

Suppose $x \in M-k(M)$. There exists a component S of $X-M$ such that

$x=F(S)$. If R denotes the sum of all the remaining components of $X-M$, it is easy to show by 1.9 and 1.10 that

$$S\cdot[\overline{(M-x)+R}]=0=\overline{S}\cdot[(M-x)+R]$$

and since $X-x=[(M-x)+R]+S$, the lemma follows.

2.20. Corollary. *If M is a true cyclic element and S is any component of $X-M$, then the point $x=F(S)$ separates any point in S from every point in $M-x$.*

2.21. Lemma. *If E is any closed set, then*

$$F(\sum S)=\overline{\sum F(S)},$$

where the summation extends over any class of components S of $X-E$.

Proof. Clearly $\sum F(S)\subset F(\sum S)$, and as the frontier of any set is closed (see 1.5),

$$\overline{\sum F(S)}\subset F(\sum S).$$

Suppose $p\in F(\sum S)$ and $p\in X-\overline{\sum F(S)}$. Consider U, the component of $X-\overline{\sum F(S)}$ containing p. For some term S of $\sum S$, $U\cdot S\neq 0$. Now $U=U\cdot S+(U-S)$. Since $U\cdot S$ is open (1.8), $(U\cdot S)\cdot(\overline{U-S})=0$. Also $(U-S)\cdot(\overline{U\cdot S})=(U-S)\cdot[U\cdot S+F(U\cdot S)]=(U-S)\cdot F(U\cdot S)\subset(U-S)\cdot[F(U)+F(S)]=(U-S)\cdot F(S)=0$. This contradicts the connectedness of the set U (1.6). Thus $F(\sum S)\subset\overline{\sum F(S)}$.

2.22. If M is a true cyclic element and if E is a subset of M, then E^* will denote the set $E+\sum S$ where the summation is taken over all components S of $X-M$ such that $F(S)\in E$.

2.23. Lemma. *If M is a true cyclic element and the set E is closed in M, then the E^* is closed in X.*

Proof.

$$\begin{aligned}\overline{E^*}&=\overline{E+\sum S}=\overline{E}+\overline{\sum S}=E+\sum S+F(\sum S)\\&=E+\overline{\sum F(S)}+\sum S=E+\sum S=E^*.\end{aligned}$$

The various equalities follow easily from 2.21 and the fact that E is closed in X.

2.24. Corollary. *If M is a true cyclic element and the set E is open in M, then E^* is open in X.*

2.25. Lemma. *If the set E is connected in M, then E^* is connected in X.*

The lemma follows at once from the fact that if E is connected in M it is connected in X by 1.6.

2.26. Lemma. *If A and B are subsets of some true cyclic element M and $A\cdot\bar{B}=0$, then $A^*\bar{B}^*=0$.*

Proof. If $\sum_A S$ and $\sum_B S$ denote the sums of all components of $X-M$ whose frontiers are in A and B respectively, then by 2.21

$$\begin{aligned}
A^*\bar{B}^* &= \Big(A+\sum_A S\Big)\cdot\Big(\overline{B+\sum_B S}\Big)\\
&= \Big(A+\sum_A S\Big)\cdot\Big[\Big(B+\sum_B S\Big)+F\Big(B+\sum_B S\Big)\Big]\\
&= \Big(A+\sum_A S\Big)\cdot F\Big(B+\sum_B S\Big)\\
&\subset \Big(A+\sum_A S\Big)\cdot\Big[F(B)+F\Big(\sum_B S\Big)\Big]\\
&= \Big(A+\sum_A S\Big)\cdot\Big[F(B)+\overline{\sum_B F(S)}\Big]\\
&\subset \Big(A+\sum_A S\Big)\cdot\bar{B}\subset\Big(\sum_A S\Big)\cdot\bar{B}=0,
\end{aligned}$$

since $A\bar{B}=0$ and $\bar{B}\subset M\subset X-\sum_A S$.

2.27. Theorem. *If Z is a connected set and M is a true cyclic element, then MZ is connected.*

Proof. If MZ is degenerate, the theorem is obvious (1.6). Suppose $MZ=A+B$, where $A\neq 0\neq B$ and $A\bar{B}=0=\bar{A}B$. Now if $p\in Z-M$, suppose that S is the component of $X-M$ that contains p. Clearly $F(S)\subset MZ$, for otherwise $Z=(SZ)+(Z-S)$ would be a decomposition showing that Z is not connected. Thus $p\in(MZ)^*$. Hence $Z\subset(MZ)^*=A^*+B^*$, and $A^*\bar{B}^*=0=\bar{A}^*B^*$ by 2.26. On the other hand $ZA^*\neq 0\neq ZB^*$; hence the connectedness of Z is contradicted.

The validity of this theorem seems surprising in view of the fact that the proof given for peano spaces (Kuratowski and Whyburn [4]) depended upon the metric and the fact that the number of components of an open set is at most denumerable. The consequences of this product theorem are as varied and important here as in peano spaces.

2.28. Corollary. *If x is a point such that $\bar{x}=x$ and M is a true cyclic element, then the set $M-x$ is connected.*

The truth of this corollary follows from 2.9 and 2.27.

2.29. Theorem. *A true cyclic element is a connected, locally connected, topological space.*

Proof. The connectedness follows from 2.27 and the fact that X is connected. Let G be open in a cyclic element M. It will be shown that each component of G is open in M. By 2.24 the set G^* is open in X. Let S^* be any component of G^*. By 2.27, S^*M is connected and nonvacuous. Using 2.25, S^*M is a component of G. But S^* is open in X by 1.8; hence S^*M is open in M. Using 2.25 it is easy to see that all the components of G are obtained in this manner.

Finally the subsets of M which are open in M clearly satisfy A1, A2 and A3 of 1.1.

III. The hyperspace

3.1. Paralleling the notation in the space X, the symbol X_h will be reserved for the abstract set composed of all cyclic elements (both true and degenerate) of the space X. Elements of the set X_h will be denoted by small Greek letters; sets of elements in X_h will be denoted by large Greek letters.

3.2. It follows easily from 2.12 and 2.17 that, if x is any point of the space X, there exists a smallest cyclic element in X which contains the point x. This fact makes possible the definition of a single-valued transformation $T(X)=X_h$ from any locally connected topological space X to the class X_h of its cyclic elements as follows: *if x is any point of the space X, then $T(x)=\xi$, where the element ξ of X_h is the smallest cyclic element of X which contains the point x.*

The following familiar conventions will be employed.

If A is a subset of X, then $T(A)=E_\xi[\xi=T(x),\ x\in A]$ will be called the *image* of the set A.

If Δ is a subset of X_h, then $T^{-1}(\Delta)=E_x[T(x)\in\Delta]$ will be called the *inverse* of Δ in X. A set D in X will be called an *inverse set* if and only if there is some set Δ such that $T^{-1}(\Delta)=D$.

It is clear that the inverse of a single element of X_h is either a degenerate cyclic element in X or the kernel of some true cyclic element in X (see 2.15).

3.3. The transformation, T, will now be used to topologize the set X_h in accordance with the convention: *a subset Γ of X_h will be called open if and only if its inverse, $T^{-1}(\Gamma)$, is an open set in the space X.*

Theorem. *The class of open sets in X_h satisfies the axioms* A1, A2, *and* A3.

Proof. The theorem follows at once from the formulas $T^{-1}(\Gamma_1\cdot\Gamma_2)=T^{-1}(\Gamma_1)\cdot T^{-1}(\Gamma_2)$, and $T^{-1}(\Sigma\Gamma_\alpha)=\Sigma T^{-1}(\Gamma_\alpha)$.

The set X_h may now be thought of as a topological space. As such, it will be termed the hyperspace of X.

3.4. A transformation from one topological space to another is said to be *continuous* if the inverse of an open set is open. The transformation is said to be *strongly continuous* if, in addition, any open inverse set has an open image (Alexandroff and Hopf [1, p. 65]).

3.5. THEOREM. *The transformation $T(X)=X_k$ is strongly continuous under the topologization of X_k given in 3.2, and this is the only topologization which will make T strongly continuous.*

3.6. LEMMA. *A set Φ is closed in X_k if and only if the set $T^{-1}(\Phi)$ is closed in X.*

3.7. LEMMA. *If E is connected in X then $T(E)$ is connected in X_k.*

The first of these lemmas is obvious; the second is well known.

3.8. It will be shown that the hyperspace X_k is a locally connected topological space. The proof requires the

LEMMA. *If Δ is any set in X_k and Σ is a component of Δ, then $T^{-1}(\Sigma)$ is the sum of certain components of the set $T^{-1}(\Delta)$.*

Proof. Let $x \in T^{-1}(\Sigma)$ and let S denote the component of $T^{-1}(\Delta)$ which contains x. Now $T(S)$ is connected by 3.7 and intersects Σ. Thus $T(S) \subset \Sigma$. Hence $S \subset T^{-1}(\Sigma)$ and the lemma follows.

THEOREM. *The space X_k is locally connected.*

Proof. If Γ is open in X_k and Σ is a component of Γ, then $T^{-1}(\Sigma)$ is the sum of certain components of $T^{-1}(\Gamma)$. But, as T is continuous, the set $T^{-1}(\Gamma)$ is open; since X is locally connected, each component of $T^{-1}(\Gamma)$ is open. Therefore $T^{-1}(\Sigma)$ is open by A3. But $T^{-1}(\Sigma)$ is an inverse set so $\Sigma = T(T^{-1}(\Sigma))$ is open by the strong continuity of T.

3.9. Consider the results achieved so far. If X is any locally connected topological space, then by 2.29, so is each true cyclic element of X; and, by 3.3 and 3.8, so is the hyperspace X_k. Thus a type of permanence of form is exhibited by the class of all locally connected topological spaces. This remark will furnish the basis of the discussion of the final section of the paper (§IV). It has also been shown that the hyperspace X_k is always related to the original space X by a strongly continuous, single-valued transformation, and that, relative to this property, the topology in X_k is uniquely determined.

To complete the solution of the problem proposed in 0.1 it remains only to show that the hyperspace X_k is always acyclic. The proof of this result seems to be difficult. It will be accomplished through a sequence of lemmas of which the first four are concerned with the components of certain sets in X_k, and the remainder deal with the relationship of conjugacy in X_k.

3.10. LEMMA. *A subset E of the space X is an inverse set if and only if, M being any true cyclic element, $E \cdot k(M) \neq 0$ implies $E \supset k(M)$.*

The proof is obvious (3.2).

3.11. LEMMA. *If E is an inverse set of X such that for any true cyclic element M whose kernel is not in E the set $E \cdot M$ is degenerate, then any component*

S of $X-E$ is an inverse set, and if E is closed, then $T(S)$ is a component of $X_h-T(E)$.

Proof. Consider any component S of $X-E$. Let M be any true cyclic element such that $k(M)\cdot S\neq 0$. Then $E\cdot k(M)=0$ by 3.10. If $E\cdot M=0$, then $S\supset M$ as M is connected (2.29). If $E\cdot M=x$, then $x\notin k(M)$ and $\bar{x}=x$ (2.19). Hence $M-x$ is connected (2.28), and lies in $X-E$. Therefore, $k(M)\subset M-x\subset S$. Hence, by 3.10, the set S is an inverse set.

Now suppose E is closed. Let R denote the sum of all the components of $X-E$ distinct from S. As above, R is a sum of inverse sets and so is an inverse set. Moreover, S and R are open; hence $T(S)$ and $T(R)$ are open and have no points in common (3.2). Therefore, $T(S)\cdot\overline{T(R)}=0=\overline{T(S)}\cdot T(R)$ by 1.9. Now $T(S)\subset X_h-T(E)$ and is connected (3.7). Therefore $T(S)$ is a component of $X_h-T(E)$.

COROLLARY. *If M is a true cyclic element and S is a component of $X-M$, then S is an inverse set and $T(S)$ is a component of $X_h-T(M)$.*

3.12. LEMMA. *With the hypotheses of 3.11, every component Σ of $X_h-T(E)$ is the image of some component S of $X-E$.*

Proof. The set $T^{-1}(\Sigma)$ is a sum of certain components of $X-E$ (3.8). By 3.11, each of these components is an inverse set and their images are components of $X_h-T(E)$. However, these images all intersect Σ and therefore are identical to Σ.

3.13. LEMMA. *If E is a subset of a true cyclic element M and S is a component of $X-M$ whose frontier is in E, then S is a component of $X-E$.*

Proof. Consider any $A\supset S$ such that $A\subset X-E$. Let $B=A-S$. Now $\bar{B}\cdot S=0$ and $\bar{S}\cdot B=[S+F(S)]\cdot B=B\cdot F(S)=0$, as $F(S)\in E$ and $B\subset X-E$. Hence S is a component of $X-E$.

3.14. LEMMA. *If M is a true cyclic element in X and if $a\in M-k(M)$ and $\alpha=T(a)$, $\beta=T(k(m))$; then $\alpha\sim\beta$.*

Proof. Suppose that α is not conjugate to β; then there is some point $\xi=\bar{\xi}$ in X_h such that ξ separates α from β. Let $b\in k(M)$. Then $a+b\subset X-T^{-1}(\xi)$ and the components of $X-T^{-1}(\xi)$ that contain a and b, respectively, must be distinct (3.7). Now $T^{-1}(\xi)$ cannot be a single point for if it were it would separate a and b, denying the fact that $a\sim b$ as M is coherent.

Then $T^{-1}(\xi)=k(M_0)$ for some true cyclic element M_0, and $M\neq M_0$ since $\xi\neq\beta$ and $T^{-1}(\beta)=k(M)$. Thus $a+b\subset M\subset X-T^{-1}(\xi)$. But M is connected (2.29) and this contradicts the fact that the components of $X-T^{-1}(\xi)$ which contain a and b are distinct.

3.15. LEMMA. *If α and β are distinct conjugate points in X_h, then either* (1)

one of the sets $T^{-1}(\alpha)$, $T^{-1}(\beta)$ *is the kernel*, $k(M)$, *of some true cyclic element* M *and the other is a single point in* $M-k(M)$; *or* (2) *both* $T^{-1}(\alpha)$ *and* $T^{-1}(\beta)$ *are single points which are conjugate.*

Proof. The proof is left to the reader. It follows easily from the fact that if the two inverse sets $T^{-1}(\alpha)$ and $T^{-1}(\beta)$ are separated in X by a point x (which is a closed inverse set in X), then α is separated from β by $T(x)$ in X_h (see 3.11).

3.16. LEMMA. *If* Λ *is an N-set in* X_h *such that the inverse of every point* $\xi\in\Lambda$ *is degenerate, then* $T^{-1}(\Lambda)=N$ *is an N-set in* X *and* $k(N)=T^{-1}[k(\Lambda)]$.

Proof. Let α and β be distinct points in Λ whose inverses are a and b respectively. By 3.15, $a\sim b$. It follows by 2.11 that there exists a unique N-set, $N\subset X$, which contains $a+b$. By the coherence and completeness of N, $T^{-1}(\Lambda)\subset N$.

If y is an arbitrary point of N, it will be shown that the image $T(y)=\eta$ is in Λ. If this were false, then by 2.10 one of $\alpha\sim\eta$ or $\beta\sim\eta$ would be false. If some point ξ in X_h separates α from η, then $T^{-1}(\xi)$ is closed; moreover, if S_a and S_y denote the components of $X-T^{-1}(\xi)$ containing a and y, respectively, $S_a\cdot S_y=0$. Otherwise, $S_a=S_y$. Then, by 3.7, $T(S_a)$ is connected, contained in $X_h-\xi$, and contains $\alpha+\eta$. This contradicts the separation of α and η by the point ξ. Since $a\sim y$, it follows by 2.2 that $T^{-1}(\xi)$ is nondegenerate. But then $T^{-1}(\xi)=k(M)$ for some true cyclic element M. Three cases arise: (1) $N\cdot M=0$, (2) $N\cdot M$ is a single point x, and (3) $N\equiv M$ (see 2.12 and 2.15).

Cases (1) *and* (2). In either case the set $N-N\cdot M$ is contained in a single component S of $X-N\cdot M$ (2.9). Thus $a+y\subset N\subset\bar{S}\subset X-k(M)$. But $T(\bar{S})$ is connected (3.7) and contained in $X_h-T[k(M)]=X_h-\xi$. This contradicts the separation of α and η by the point ξ.

Case (3). If $N\equiv M$, then $T^{-1}(\xi)=k(M)$ and 3.15 together imply that $T^{-1}(\alpha)=a\in M-k(M)$. Hence, by 3.14, $\xi\sim\alpha$. Similarly, $\xi\sim\beta$. But then $\xi\in\Lambda$ and $T^{-1}(\xi)$ is nondegenerate which contradicts the hypothesis.

It has been established that $T^{-1}(\Lambda)=N$. It remains to show that $T^{-1}[k(\Lambda)]=k(N)$.

Let ξ be any point in $k(\Lambda)$ and set $x=T^{-1}(\xi)$. If $x\notin k(N)$, then for some component S of $X-N$, $x\in F(S)$. Since N fulfills the hypotheses of the set E of 3.11, the component S is an inverse set and $T(S)$ is a component of $X_h-T(N)=X_h-\Lambda$. If Γ is any open set containing $\xi=T(x)$, the set $T^{-1}(\Gamma)$ is open and contains x. But then $x\in F(S)$ implies $\Gamma\cdot T(S)\neq 0$ and thus $\xi\in F[T(S)]$ contrary to $\xi\in k(\Lambda)$. Thus $x\in k(N)$, and $T^{-1}[k(\Lambda)]\subset k(N)$.

Finally, $k(N)\subset T^{-1}[k(\Lambda)]$. To see this, let $x\in k(N)$ and suppose that $T(x)=\xi\notin k(\Lambda)$. Since $\xi\in\Lambda$, there is some component Σ of $X_h-\Lambda$ such that $\xi\in F(\Sigma)$. By 3.12, $T^{-1}(\Sigma)$ is a component of $X-N$. But $x\in k(N)$ implies $x\notin F[T^{-1}(\Sigma)]$; therefore $x\notin\overline{T^{-1}(\Sigma)}$. Moreover, by 3.10, $\overline{T^{-1}(\Sigma)}$ is an inverse

set so $X-\overline{T^{-1}(\Sigma)}$ is an open inverse set containing x. Thus the open set $T[X-\overline{T^{-1}(\Sigma)}]$ contains the point ξ and is disjoint of Σ; this contradicts $\xi\in F(\Sigma)$.

The lemma follows.

3.17. THEOREM. *The hyperspace X_h contains no true cyclic elements.*

Proof. Consider any N-set, $\Lambda\subset X_h$. There are two cases.

Case (1). If there exists a point $\alpha\in\Lambda$ such that $T^{-1}(\alpha)$ is nondegenerate, it will be shown that $\alpha\supset k(\Lambda)$ and so Λ cannot be a true cyclic element (2.15). Now $T^{-1}(\alpha)=k(M)$ for some true cyclic element M. By 3.15 for any point $\xi\neq\alpha$ contained in Λ, $T^{-1}(\xi)$ is a point $x=F(S)$ for some component S of $X-M$. By 3.11, $T(S)$ is a component of $X_h-\Lambda$ and $\xi\in F[T(S)]$. Thus $\alpha\supset k(\Lambda)$.

Case (2). For every point $\alpha\in\Lambda$, $T^{-1}(\alpha)$ is degenerate. By 3.16, $T^{-1}(\Lambda)=N$ is an N-set and $T^{-1}[k(\Lambda)]=k(N)$. If N is a true cyclic element in X or if $k(N)$ is degenerate, then $k(\Lambda)$ is a single point. In these cases Λ is not a true cyclic element (2.15). If N is not a true cyclic element but $k(N)$ is nondegenerate, then for some component S of $X-N$ the frontier $F(S)$ contains x and y where $x\neq y$. By 3.11, the set $T(S)=\Sigma$ is a component of $X_h-\Lambda$. Also, x and y are distinct inverse sets. This implies that $F(\Sigma)$ is nondegenerate. By 2.15, Λ is not a true cyclic element and the theorem is established.

3.18. In 2.1 the space X was assumed to be connected. That this restriction imposed no loss of generality is easy to see. If the space X is not connected, the definitions in 2.2 and 2.3 of separation and conjugacy need only be regarded relative to the individual components X^* of the space X. Thus, a point p will be said to separate two points a and b if (1) the points a, b, and p lie in a single component X^* of X, (2) $\bar{p}=p$, and (3) a and b lie in distinct components of X^*-p. Similarly, two points a and b are conjugate if they lie in the same component X^* of X and are separated by no point of X^*.

The hyperspace X_h of X will be the totality of cyclic elements of all the components of X, each being topologized according to 3.2.

A locally connected topological space is said to be *acyclic* if its hyperspace is topologically equivalent to itself. If a space has no true cyclic elements, then it is easily seen to be acyclic; conversely, if a space is acyclic, it has no true cyclic elements.

THEOREM. *If X is a locally connected topological space, then each true cyclic element of X and the hyperspace X_h are locally connected topological spaces. Moreover, the hyperspace X_h is acyclic.*

Proof. The proof is a direct consequence of 2.29, 3.8, 3.17, and the remarks above.

IV. Hereditary classes of spaces

4.1. The theorem of 3.18 suggests the following definition. A subclass $\mathfrak{H}$ of the class $\mathfrak{X}$ of all locally connected topological spaces will be called *hereditary* if whenever X is in $\mathfrak{H}$, the hyperspace X_h is in $\mathfrak{H}$ and each true cyclic element of X is a member of $\mathfrak{H}$. The theorem of 3.18 can now be phrased as:

The totality of locally connected topological spaces is a hereditary class of spaces.

Remark. Any hereditary class is cyclicly reducible (Kuratowski and Whyburn [4]).

Consider the class $\mathfrak{P}$ of all peano spaces. The example consisting of two tangent circles in the plane is a member of the class. However, the hyperspace for this example contains only three points. Since such a space cannot be connected and metric, it is not a peano space. Thus, the class $\mathfrak{P}$ is not a hereditary class. On the other hand, every peano space is a locally connected topological space; therefore there exists a hereditary class which contains $\mathfrak{P}$. An immediate question is: *what is the smallest class* (if it exists) *which is hereditary and which contains all peano spaces?* This question is answered in 4.3 and 4.4.

4.2. Theorem. *The logical product or sum of any number of hereditary classes is a hereditary class.*

The proof is immediate from the definitions.

4.3. Let $\mathfrak{P}^*$ be the product of all the hereditary classes which contain the class $\mathfrak{P}$ of all peano spaces. One immediately obtains the

Theorem. *The class $\mathfrak{P}^*$ is the smallest hereditary class which contains all* peano spaces.

This result demonstrates the existence of an answer to the question posed in 4.1. A constructive definition of the class $\mathfrak{P}^*$ is found in the following section.

4.4. Theorem. *A necessary and sufficient condition that a space X be contained in the class $\mathfrak{P}^*$ is that X be either a peano space or the hyperspace of a peano space.*

Proof. Let $\mathfrak{P}$ and $h[\mathfrak{P}]$ denote, respectively, the class of all peano spaces and the class of all their hyperspaces. The class $\mathfrak{P}+h[\mathfrak{P}]$ is hereditary by 3.18 and because every true cyclic element of a peano space is in $\mathfrak{P}$. Then by the definition of $\mathfrak{P}^*$ one has $\mathfrak{P}^*\subset\mathfrak{P}+h[\mathfrak{P}]$. The reverse inclusion follows since any hereditary class that contains $\mathfrak{P}$ must contain $h[\mathfrak{P}]$.

4.5. The class $\mathfrak{P}$ may be defined as the totality of compact, metric spaces which are connected and locally connected. It would be interesting to know

an analogous intrinsic definition for the class $\mathcal{P}^*$. With this in view, the remainder of the paper is devoted to a list of intrinsic properties of $\mathcal{P}^*$.

4.6. A subclass of $\mathfrak{X}$ is *invariant* with respect to strongly continuous transformations if every strongly continuous image of a member of the class is again a member of the class (3.4).

THEOREM. *Any invariant subclass of $\mathfrak{X}$ is hereditary.*

Proof. As the hyperspace X_h of a space X is a strongly continuous image of X (see 3.5), half of the theorem is proved. For the last part we need the

THEOREM. *A true cyclic element of X is a strongly continuous image of X.*

Proof. Let M denote any true cyclic element in the space X. Define the transformation $f(X)=M$ by

$$f(x) = \begin{cases} x \text{ if the point } x \text{ is in } M, \\ F(S) \text{ if the point } x \text{ is in the component } S \text{ of } X - M. \end{cases}$$

Let A be any open set in M; then $f^{-1}(A)=A^*$; and A^* is open (2.22 and 2.24). On the other hand if A is a set in M such that $f^{-1}(A)=A^*$ is open in X, then A is clearly open in M because $A=M\cdot A^*$. Thus $f(X)=M$ is strongly continuous. (M is, in fact, a retract, see Borsuk [10].)

4.7. The reader will have no difficulty in proving that these subclasses of $\mathfrak{X}$ are invariant and hence hereditary (4.6):

(1) the class of connected spaces,

(2) the class of separable spaces,

(3) the class of compact spaces (for any of the usual interpretations of compactness),

(4) the class of spaces having the property that the number of components of an open set is denumerable,

(5) the class of spaces which are strongly continuous images of some fixed space.

An interesting subcase of (5) occurs when the *fixed space is the closed unit interval.* In this case the class is a subclass of the four preceding classes.

A tabulation of hereditary classes might be continued on the above lines. However, there are hereditary classes which are not invariant and some of these will be considered in the concluding sections of the paper.

4.8. Before turning to this topic some remarks should be made about the following separation axiom, attributed to Kolmogoroff (Alexandroff and Hopf [1, p. 58]).

T_0: *If x and y are distinct points, then at least one of them is contained in an open set that does not contain the other.*

This axiom will be referred to as the T_0-property and topological spaces in which it is satisfied will be called T_0-spaces.

If x and y are distinct points of a T_0-space such that $x \in \bar{y}$, then x will be called a *satellite* of y.

Clearly, if x is a satellite of y, then y cannot be a satellite of x. This follows from the hypothesis $x \in \bar{y}$ and the T_0-property which then requires $y \cdot \bar{x} = 0$. Thus one has the

THEOREM. *In a T_0-space, the relation 'x is a satellite of y' is asymmetric.*

4.9. It will be shown in 4.10 that the class of locally connected T_0-spaces is hereditary. This result requires the

LEMMA. *If x is a point in any locally connected topological space, then the set $\bar{x}$ is coherent* (*see* 2.7.)

Proof. The lemma is trivial if $\bar{x} = x$. Suppose that for $y \neq x$, it is true that $y \in \bar{x}$, and z is a point distinct from x and y and such that $\bar{z} = z$. If S denotes the component of $X - z$ containing x, then $x + y$ is contained in S since $x + y$ is connected (1.6). Thus $y \sim x$. The lemma follows by 2.5.

COROLLARY. *If a point x in any locally connected topological space is an inverse set, then so is $\bar{x}$ an inverse set* (*see* 3.2.)

Proof. Let M be any true cyclic element and suppose that $\bar{x} \cdot k(M) \neq 0$. It follows by the coherence of $\bar{x}$ and the corollary of 2.20 that $x \in M$. But x is an inverse set so $x \in M - k(M)$. Thus, by 2.19, it is true that $\bar{x} = x$, and the theorem follows. If $\bar{x} \cdot k(M) = 0$, then $\bar{x}$ is an inverse set by 3.10.

4.10. THEOREM. *The class of all locally connected T_0-spaces is hereditary.*

Proof. Let X be any space in the class $\mathfrak{X}$; suppose the T_0-property satisfied in X. Let ξ and η be distinct points in the hyperspace X_k. There are two cases according as (1) at least one of the inverses $T^{-1}(\xi)$ and $T^{-1}(\eta)$ is nondegenerate, or (2) both inverses are degenerate.

Case 1. Suppose $T^{-1}(\xi) = k(M)$ for some true cyclic element M. If $M \cdot T^{-1}(\eta) = 0$, let S be the component of $X - M$ which contains $T^{-1}(\eta)$; S exists by 2.29. Now S is an open inverse set so $T(S)$ is an open set containing η but not ξ (3.11). If $M \cdot T^{-1}(\eta) \neq 0$, then $T^{-1}(\eta) = y \in M - k(M)$ by 2.12 and 2.15. Let S be the component of $X - y$ containing $k(M)$ (2.28). Since $\bar{y} = y$, it follows that S is open and $T(S)$ is an open set containing ξ but not η (3.11).

Case 2. Let $x = T^{-1}(\xi)$ and $y = T^{-1}(\eta)$. By the T_0-property in x, either $y \notin \bar{x}$ or $x \notin \bar{y}$; assume the former. By 4.9, $\bar{x}$ is an inverse set. Hence 3.11 may be applied as in Case 1 to obtain an open set containing η but not ξ.

Remark. The example of 4.1 shows that the class of all locally connected T_1-spaces is not hereditary (Alexandroff and Hopf [1, p. 58]).

4.11. The hereditary classes discussed so far have been of a general nature.

In those to follow, attention is focused more explicitly on certain properties of peano spaces.

If M is any true cyclic element in a peano space, then the kernel of M is such that its closure contains M. Let $\mathfrak{K}_1$ be the totality of spaces in $\mathfrak{X}$ for which $\overline{k(N)}=N$ for any N-set.

THEOREM. *The class $\mathfrak{K}_1$ is hereditary.*

Proof. Let X be any space in the class $\mathfrak{K}_1$ and let Λ be any N-set in X_h. There are two cases.

Case 1. If every point in Λ has a degenerate inverse, the result follows from 3.16 and the hypothesis that $\overline{k(N)}=N$ in X.

Case 2. Suppose that some point $\xi\in\Lambda$ has a nondegenerate inverse $k(M)$, where M is a true cyclic element in X. By 3.15, $T^{-1}(\Lambda)\subset M$. By the continuity of T and the hypothesis, $\overline{k(N)}=N$, it is true that $T(M)\subset\Lambda$. Hence M is the inverse of Λ. It follows (3.11) that $k(\Lambda)=\xi$ and $\bar{\xi}=\Lambda$.

4.12. LEMMA. *If X is in $\mathfrak{K}_1$, then for every point ξ in X_h, one has $T^{-1}(\bar{\xi})=\overline{T^{-1}(\xi)}$.*

Proof. Clearly $T^{-1}(\bar{\xi})\supset T^{-1}(\xi)$ and hence

$$T^{-1}(\bar{\xi}) = \overline{T^{-1}(\bar{\xi})} \supset \overline{T^{-1}(\xi)}.$$

To see the reverse inclusion, consider first the case in which $T^{-1}(\xi)=x$, a single point. Since x is an inverse set, the same is true of $\bar{x}$ (4.9). Thus if any point y is not in $\bar{x}=\overline{T^{-1}(\xi)}$, then y is in some component S of $X-\bar{x}$. By the corollary of 4.9, $\bar{x}$ satisfies the hypotheses of 3.11. Hence S is an inverse set and $T(S)$ is a component of $X_h-T(\bar{x})\subset X_h-\xi$. Thus there is an open set, $T(S)$, containing $T(y)$ but not ξ. Hence $T(y)\notin\bar{\xi}$. Therefore $T^{-1}(\bar{\xi})\subset\overline{T^{-1}(\xi)}$. In case $T^{-1}(\xi)=k(M)$ for some true cyclic element M, it follows that $T^{-1}(\bar{\xi})=M=\overline{k(M)}=\overline{T^{-1}(\xi)}$ and the lemma is proved.

4.13. COROLLARY. *Under the hypotheses of the lemma, if ξ and η are distinct points in the hyperspace X_h of X, then $\xi\in\bar{\eta}$ if and only if either*: (1) *$T^{-1}(\xi)$ and $T^{-1}(\eta)$ are single points such that $T^{-1}(\xi)\in\overline{T^{-1}(\eta)}$, or* (2) *$T^{-1}(\eta)=k(M)$ for some true cyclic element M, and $T^{-1}(\xi)=x\in M-k(M)$.*

The proof is easy and will be left to the reader.

4.14. Let $\mathfrak{K}_2$ be the class of all spaces X in $\mathfrak{K}_1$ having the property that if x is any point, the frontier of any component of $X-\bar{x}$ is a single point.

THEOREM. *The class $\mathfrak{K}_2$ is hereditary.*

Proof. Consider $X\in\mathfrak{K}_2$. If M is a true cyclic element in X it certainly has the property (2.25). Let ξ be any point in X_h. If $\bar{\xi}=\xi$, the result is im-

mediate. Suppose $\bar{\xi} \neq \xi$. By 4.12, $T^{-1}(\bar{\xi}) = \overline{T^{-1}(\xi)}$. If $T^{-1}(\xi)$ is degenerate, then any component of $X - \overline{T^{-1}(\xi)}$ has a single frontier point by hypothesis. If $T^{-1}(\xi) = k(M)$ for some true cyclic element, then $M = \overline{T^{-1}(\xi)}$ and again any component of $X - \overline{T^{-1}(\xi)}$ has a single frontier point. It follows easily with the help of 3.11 and 3.12 that every component of $X_h - \bar{\xi}$ has a single frontier point.

4.15. Let $\mathfrak{K}_3$ be the totality of spaces X in $\mathfrak{K}_2$ such that if N is any N-set, then the frontier of any component of $X - N$ is a single point.

THEOREM. *The class $\mathfrak{K}_3$ is hereditary.*

Proof. Suppose $X \in \mathfrak{K}_3$. Let Λ be any N-set in X_h, and let Σ be any component of $X_h - \Lambda$. If for some point $\xi \in \Lambda$, it is true that $T^{-1}(\xi) = k(M)$ for some true cyclic element M, then $\overline{T^{-1}(\xi)} = M$ and $\bar{\xi} = \Lambda$. The frontier $F(\Sigma)$ is a single point by 4.14. Suppose that for every point $\xi \in \Lambda$ the set $T^{-1}(\xi)$ is degenerate. By 3.16, $T^{-1}(\Lambda) = N$ is an N-set and $k(N) = T^{-1}[k(\Lambda)]$. By the assumption on the inverses of points in Λ, the set N is not a true cyclic element. It follows by 2.15 that the kernel $k(N)$ is degenerate. But $\overline{k(N)} = N$ and hence N is of the form $\bar{x}$ where x is a single point. Then $F(\Sigma)$ is a single point by 4.14.

4.16. THEOREM. *Let X be in $\mathfrak{K}_2$. If x is any point in X such that $\bar{x} \neq x$, then the set $\bar{x}$ is an N-set.*

Proof. The set $\bar{x}$ was proved coherent in 4.9. Let y be any point in $X - \bar{x}$. Let S denote the component of $X - \bar{x}$ that contains y and let R denote the sum of the remaining components of $X - \bar{x}$. As in the proof of 2.19, one shows that the point $F(S)$ separates y from every point in $\bar{x} - y$. It follows that $\bar{x}$ is complete and the theorem is proved.

COROLLARY. *Let X be any T_0-space in $\mathfrak{K}_2$. If the point $x \in X$ is a satellite of the point y, then y is not a satellite of any point.*

Proof. If $x \in \bar{y}$ and $y \in \bar{z}$, then $z \neq x$ since $y \cdot \bar{x} = 0$. Moreover, $z \cdot \bar{y} = 0$. Let S be the component of $X - \bar{y}$ that contains z. Now $\bar{z} \subset \bar{S} = S + F(S)$ and $F(S)$ is a single closed point. Thus $F(S) \neq y$ since $\bar{y} \neq y$. But then $y \cdot \bar{z} = 0$ which is contradictory.

4.17. **Conclusion.** Let Υ denote the totality of all locally connected topological spaces which are (1) strongly continuous images of the closed unit line interval, (2) T_0-spaces in $\mathfrak{K}_3$. By the theorem of 4.2, Υ is a hereditary class. Moreover, it contains the class $\mathcal{P}^*$ (see 4.3). More restrictive hereditary classes which contain $\mathcal{P}^*$ will suggest themselves to the reader. The class Υ has been mentioned because it yields considerable insight into the nature of the hyperspaces of peano spaces.

It is a consequence of 2.15 and 4.15 that in any space of the class Υ every N-set is either a true cyclic element or a set of the form $\bar{x}$ where x is a point. Thus, in any peano hyperspace, the only N-sets are of the form $\bar{\xi}$ where ξ is

a point. The corollary of 4.13 states explicitly how such N-sets arise; namely, from the incidence in the original peano space of a degenerate cyclic element on a true cyclic element. This relationship is certainly not symmetrical. Therefore, the asymmetry of the T_0-property seems particularly fitting in this connection.

BIBLIOGRAPHY

1. P. Alexandroff and H. Hopf, *Topologie* I, Berlin, 1935.

2. F. Hausdorff, *Mengenlehre*, Berlin, 1927 or 1935.

3. B. Knaster and C. Kuratowski, *Sur les ensembles connexes*, Fundamenta Mathematicae, vol. 2 (1921), p. 206.

4. C. Kuratowski and G. T. Whyburn, *Sur les éléments cycliques et leurs applications*, Fundamenta Mathematicae, vol. 16 (1930), p. 305.

5. G. T. Whyburn, *Cyclicly connected continuous curves*, Proceedings of the National Academy of Sciences, vol. 13 (1927), p. 31.

6. ———, *Concerning the structure of a continuous curve*, American Journal of Mathematics, vol. 50 (1928), p. 167.

7. R. L. Moore, Rice Institute Pamphlet, vol. 23, no. 1, 1936.

8. T. Radó and P. V. Reichelderfer, *Cyclic transitivity*, Duke Mathematical Journal, vol. 6 (1940), p. 474.

9. J. W. T. Youngs, *Arc-spaces*, Duke Mathematical Journal, vol. 7 (1940), p. 68.

10. K. Borsuk, *Sur les rétractes*, Fundamenta Mathematicae, vol. 27 (1931), p. 152.

OHIO STATE UNIVERSITY,
COLUMBUS, OHIO
PURDUE UNIVERSITY,
LAFAYETTE, IND.